Omair Sarwar, MSc

Facial Privacy Protection in Airborne Recreational Videography

DISSERTATION

to gain the Joint Doctoral Degree

Doctor of Philosophy (PhD)

Alpen-Adria-Universität Klagenfurt

Fakultät für Technische Wissenschaften

in accordance with

The Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments

Alpen-Adria-Universität Klagenfurt | Queen Mary University of London

Università degli Studi di Genova

First Supervisor

Prof. Dr. Bernhard Rinner

Institute of Networked and Embedded Systems

Alpen-Adria-Universität Klagenfurt, Austria

Second Supervisor

Prof. Dr. Andrea Cavallaro

Centre for Intelligent Sensing

Queen Mary University of London, United Kingdom

October, 2018

Acknowledgments

This PhD Thesis has been developed in the framework of, and according to, the rules of the Erasmus Mundus Joint Doctorate on Interactive and Cognitive Environments EMJD ICE [FPA n° 2010-0012] with the cooperation of the following Universities:

Alpen-Adria-Universität Klagenfurt – AAU

Queen Mary, University of London – QMUL

Technische Universiteit Eindhoven – TU/e

Universitat Polittècnica de Catalunya – UPC

Università degli Studi di Genova – UNIGE

According to ICE regulations, the Italian PhD title has also been awarded by the Universittà degli Studi di Genova.

This work was supported in part by the research initiative Intelligent Vision Austria with funding from the Austrian Federal Ministry of Science, Research and Economy and the Austrian Institute of Technology.

First Reviewer

Associate Prof. Dr. Pradeep K. Atrey

Department of Computer Science

State University of New York, USA

Second Reviewer

Prof. Dr. Andreas Uhl

Department of Computer Sciences

Paris-Lodron-Universität Salzburg, Austria

Affidavit

I hereby declare in lieu of an oath that

- the submitted academic paper is entirely my own work and that no auxiliary materials have been used other than those indicated,
- I have fully disclosed all assistance received from third parties during the process of writing the paper, including any significant advice from supervisors,
- any contents taken from the works of third parties or my own works that have been included either literally or in spirit have been appropriately marked and the respective source of the information has been clearly identified with precise bibliographical references (e.g. in footnotes),
- to date, I have not submitted this paper to an examining authority either in Austria or abroad and that
- when passing on bound copies of the academic thesis, I will ensure that each bound copy is fully consistent with the submitted digital version.

I understand that the digital version of the academic thesis submitted will be used for the purpose of conducting a plagiarism assessment.
I am aware that a declaration contrary to the facts will have legal consequences.

Omair Sarwar, e.h. Klagenfurt, 31 October 2018

Dedicated to my uncle Muhammad Amanant Ali Raza, MSc (late) who was always a source of motivation and inspiration for me.

Acknowledgments

All the praise to Allah Almighty, the most magnificent and merciful. I am extremely grateful to HIM for HIS countless blessings, one of them an opportunity to study.

I am highly thankful to my supervisors, Prof. Bernhard Rinner and Prof. Andrea Cavallaro, for their immense support, guidance, and advice, without which it would not have been possible to accomplish this thesis. I would especially appreciate their valuable feedback to improve my technical writing skills.

I would take this opportunity to thank Aschbacher Heidelies for her proactive administrative support to make my life smooth during the course of my PhD studies.

Finally, I am thankful to my parents for their prayers. I would also thank my brother Aamir Nadeem Sarwar for his moral and motivational support, especially for his courage to take care of my social responsibilities back home.

Abstract

Cameras mounted on Micro Aerial Vehicles (MAVs) are increasingly used for recreational photography and videography. However, aerial photographs and videographs of public places often contain faces of bystanders thus leading to a perceived or actual violation of privacy. To address this issue, this thesis presents a novel privacy filter that adaptively blurs sensitive image regions and is robust against different privacy attacks. In particular, the thesis aims to impede face recognition from airborne cameras and explores the design space to determine when a face in an airborne image is inherently protected, that is when an individual is not recognisable. When individuals are recognisable by facial recognition algorithms, an adaptive filtering mechanism is proposed to lower the face resolution in order to preserve privacy while ensuring a minimum reduction of the fidelity of the image. Moreover, the filter's parameters are pseudo-randomly changed to make the applied protection robust against different privacy attacks. In case of videography, the filter is updated with a motion-dependent temporal smoothing to minimise flicker introduced by the pseudo-random switching of the filter's parameters, without compromising on its robustness against different privacy attacks. To evaluate the efficiency of the proposed filter, the thesis uses a state-of-the-art face recognition algorithm and synthetically generated face data with 3D geometric image transformations that mimic faces captured from an MAV at different heights and pitch angles. For the videography scenario, a small video face data set is first captured and then the proposed filter is evaluated against different privacy attacks and the quality of the resulting video using both objective measures and a subjective test.

Table of Contents

List of Symbols

A_{mz}	Amplitude of a Gaussian function
$A_{t,\Delta t}$	An activity of Δt duration starting from t
C_R	Center of a face
D	Distortion introduced by F_{Ω_j}
E	An attacker function
F_{Ω_j}	A privacy filter with selected parameter Ω_j
G_{mz}	An element of $\mathcal{G}$
H	Height of R
H_I	Height of a video frame I_t
H_s	Standard average height of a human face
I	Identity probability of an individual
I_c	Identity probability from context
I_m	Identity probability from main-identifiers
I_q	Identity probability from quasi-identifiers
I_t	A video frame at time t
I_t^p	A protecetd video frame
K	Number of subjects in $\mathcal{D}$
L	Privacy loss
M	Number of supplementary Gaussian functions
M_z	An element of $\mathcal{M}$
N	Identification rank
O	A subjective operator
P	Probability of predicting the label of a face
Q_j	Sub-region size in pixels

R	Face region in general
R_t	Face region at time t
R_{max}	Maximum intensity value of a pixel
S	Sensitivity index of a private information
T	Number of frames in a video
TN	True negatives
TP	True positives
U	Number of faces in I_t
W	Width of R
W_I	Width of a video frame I_t
W_s	Standard average width of a human face
Y	Number of images of a single subject in $\mathcal{D}$
Z	Number of sub-regions of R
ΔR	Mean square error between R and $\bar{R}$
$\bar{R}$	Protected face region
$\hat{R}$	Reconstructed face region
$\boldsymbol{N}$	Nadir direction
$\boldsymbol{P}$	Principal axis of a camera
d	Horizontal distance between the face and the camera
f	Focal length
f_s	Nuquist frequency of ρ_j
f_s^o	Nuquist frequency of ρ_i^o
h_1	Height of an MAV from ground
h_2	Height of a subject from ground
$j \in \{h, v\}$	Direction horizontal h and vertical v
p_j	Physical dimension of a pixel in j direction
Ω_j	Distortion parameters of a privacy filter (μ_j, σ_j)
Ω_j^*	Ideal distortion parameters of a privacy filter (μ_{jm}, σ_{jm})
Ω_j^o	Optimal distortion parameters of a privacy filter (μ_j^o, σ_j^o)
$\acute{\sigma_j^o}$	Frequency domain standard deviation corresponding to σ_j^o
α_s	Spatial blending weight
α_t	Temporal blending weight

α_{jm}	Pseudo-randomly generated number for μ_{jm}
$\bar{\sigma}_j$	Standard deviation of global smoothing filter
β_{jm}	Pseudo-randomly generated number for σ_{jm}
δ	Impusle response
ϵ_i	Identification accuracy of a random classifier
ϵ_v	Verification accuracy of a random classifier
η_i	Identification accuracy of a face recogniser
η_v	Verification accuracy of a face recogniser
γ_{jm}	Scaling factor for σ_j^o
$\mathcal{D}$	Data set
$\mathcal{G}$	A set of Gaussian functions
$\mathcal{K}$	A set of original identity labels
$\mathcal{M}$	A set of Gaussian mixture models
$\mathcal{R}_\mathcal{G}$	Gallery data set
$\mathcal{R}_\mathcal{P}$	Probe data set
$\mathcal{R}$	A set containing sub-regions of R
$\mathcal{X}$	Set of tuple containing parameters of Gaussian functions
$\bar{\mathcal{R}}_\mathcal{G}$	Filtered gallery data set
$\bar{\mathcal{R}}_\mathcal{P}$	Filtered probe data set
$\bar{\mathcal{R}}$	A set containing protected sub-regions of R
$\hat{\mathcal{R}}_\mathcal{G}$	Reconstructed gallery data set
$\hat{\mathcal{R}}_\mathcal{P}$	Reconstructed probe data set
ϕ	A set of weights for Gaussian mixture model
$\tilde{\mathcal{K}}$	A set of predicted identity labels
μ_j	Mean of a Gaussian PSF
μ_j^o	Mean of an optimal Gaussian PSF
μ_{jm}	Randomly modified μ_j^o for m^{th} Gaussian PSF
ν_c	Weighting factor for I_c
ν_m	Weighting factor for I_m
ν_q	Weighting factor for I_q
ω_R	Control signal for privacy design space
ϕ_{mz}	An element of ϕ

ψ_j	Convolution kernel
ρ_j	Pixel density (px/cm) in j direction
ρ_j^o	Threshold pixel density of j direction
σ_j	Standard deviation of a Gaussian PSF
σ_j^o	Standard deviation of an optimal Gaussian PSF
σ_{jm}	Randomly modified σ_j^o for m^{th} Gaussian PSF
θ_P	Tilt angle of $\boldsymbol{P}$ from $\boldsymbol{N}$
θ_R	Face capturing angle from $\boldsymbol{N}$

List of Acronyms

AC	Axis Communication
AES	Advanced Encryption Standard
AGB	Adaptive Gaussian Blur
AHGMM	Adaptive Hoping Gaussian Mixture Model
BSIA	British Security Industry Association
CCD	Charge Coupled Device
CCTV	Closed Circuit Television
CEN	European Committee for Standardization
CMOS	Complementary Metal Oxide Semiconductor
DCT	Discrete Cosine Transform
DES	Data Encryption Standard
EER	Equal Error Rate
FGB	Fixed Gaussian Blur
FIR	Far Infra-red
FV	Field of View
GARP	Gender Age Race Protector
GMM	Gaussian Mixture Model
GNN	Generative Neural Network
GPS	Global Positioning System
IF	Inverse Filter
IMU	Inertial Measurement Unit
IR	Infra-red
LBS	Location Based Service
LDA	Linear Discriminant Analysis

LED	Light Emitting Diode
LFW	Labelled Faces in the Wild
MAV	Micro Aerial Vehicle
NIR	Near Infra-red
ODBVP	Optimal Distortion-Based Visual Protector
PCA	Principal Component Analysis
PCC	Privacy using Chaos Cryptography
PDS	Privacy Design Space
PICO	Privacy through Invertible Cryptographic Obscuration
RMSE	Root Mean Square Error
PRNG	Pseudo-random Number Generator
PSF	Point Spread Function
PSNR	Peak Signal to Noise Ratio
PTC	Privacy by Trusted Computing
ROC	Receiver Operating Curve
SLM	Spatial Light Modulator
SR	Super-resolution
SSIM	Structural Similarity Index
SVGB	Space Variant Gaussian Blur
TPM	Trusted Platform Module
UAS-VPG	Unmanned Aircraft Systems-Visual Privacy Guard
3DMM	3D Morphable Model

Chapter 1

Introduction

Micro Aerial Vehicles (MAVs) are becoming common platforms for a number of civilian applications such as search and rescue [1], agriculture [2], transportation [3], news reporting [4], environmental mapping [5] and disaster management [6]. Moreover, individuals use MAVs equipped with high-resolution cameras for recreational photography and videography in public places during sports activities and social gatherings [7, 8]. Such use in public places raises privacy concerns as bystanders who happen to be within the Field of View (FV) of the camera are captured as well. Moreover, the MAV's operator can intentionally point the camera's lens wherever he/she wants (see Figure 1.1). In this thesis, operator of the MAV is called the MAV's owner, while bystander means any person other than the MAV's owner, e.g. a walking person, a vehicle-driving person, a sport-playing person or even a person inside a house but visible through its window.

The identity of a bystander is the key to his/her perceived or actual violation of privacy and it can be estimated from the main or quasi identifiers of the bystander as well as the context of the videos [9, 10]. The main-identifiers are identity sources which posses unique identifiable information of a bystander, e.g. face, ear, fingerprint and vehicle licence plate. In contrast, the quasi-identifiers are those identity sources which contain to a degree unique information of the bystander but not completely, e.g. age, race, gender, hair-style, type and colour of clothes. It has been shown that such information can assist to infer the identity of the bystander [10]. The context means location and time information that are estimated from the background of the captured videos. In this thesis, image-regions corresponding to the main-identifiers and the quasi-identifiers are called sensitive-regions. In order to preserve privacy of a bystander, these sensitive-regions are usually protected by removing, replacing

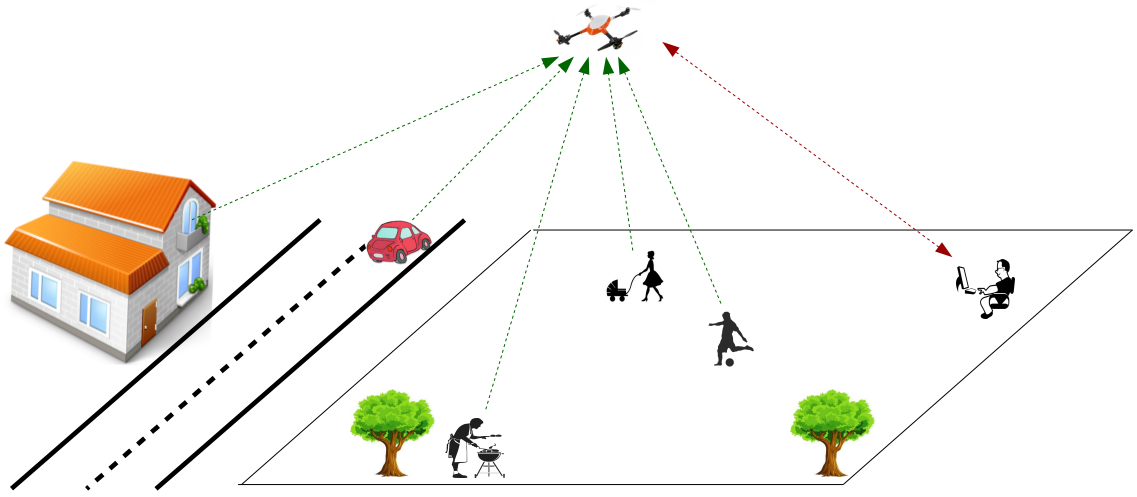


Figure 1.1: A privacy-violating scenario in an airborne-recreational videography. A camera-equipped MAV is allowed in a public place, e.g. a park, and a person (rightmost) is flying such an MAV for a recreational purpose. Although the objective of the MAV is to capture operator's video, it can also collect private information of bystanders intentionally or unintentionally, both inside and outside of the park area. Consequently, this can violate privacy of the bystanders.

or sufficiently distorting the sensitive-regions using algorithms called privacy filters [11–13]. On the other hand, the context is protected by restricting the cameras from certain locations, removing the background of the videos or just distorting it [14–16].
Unlike several applications of surveillance imagery such as people counting, perimeter protection and behaviour analysis whose utility is not compromised by a full redaction of the sensitive-regions and context [17], recreational videos require a minimal distortion of the image content in order to be usable. For these videos, the utility can be defined as the fidelity (i.e. image quality) of the protected images with respect to the originally captured images. In addition, recreational videos require automatic and robust protection of the sensitive-regions preferably on-board the MAVs as the MAV's owner cannot be ensured to be a legitimate individual, i.e. a privacy law abiding person. Automatic protection means without any involvement of the MAV's owner, while robust protection means the protected sensitive-region is not recognisable by various attacks, e.g. brute-force, naïve, parrot and reconstruction attacks [12, 13, 18–22]. A brute-force attack tries to decipher the protected probe images (i.e. sensitive-regions) by an exhaustive search [19, 20]. Other attacks use

gallery images in addition to the protected probe images [12, 13, 21, 22]. In a naïve attack, the protected probe images are compared against the unprotected gallery images [12, 13, 21]. In a parrot attack, the attacker has knowledge about the privacy filter and can transform the gallery images into the distorted domain [12]. In a reconstruction attack, the attacker has some knowledge of how to (partially) reconstruct the probe image from the protected to the unprotected domain [18]. Examples of reconstruction methods include Inverse Filter (IF) [18] and Super-resolution (SR) [22].

Thus, a privacy filter for an airborne-recreational videography should satisfy the following properties: (a) introduce only a minimal spatio-temporal distortion; (b) be robust against attacks; and (c) be computationally efficient. *Minimal spatio-temporal distortion* is necessary not to divert the attention of a viewer. *Robustness* is important to avoid privacy violations by various attacks, e.g. brute-force, naïve, parrot and reconstruction attacks. Finally, *computational efficiency* is desirable when the filter operates using the limited computational and battery power of an MAV.

1.1 Motivation and Objective

The motivation of this thesis is to develop a robust privacy filter for the main-identifiers of bystanders captured in recreational videos. The main-identifiers are the key identifiers and should be protected first compared to quasi-identifiers and context. The thesis considers the application of airborne-recreational videography using an MAV in a public place. This motivation requires an investigation of the existing privacy filters intended for the main-identifiers especially those developed for the video surveillance, i.e. Closed Circuit Television (CCTV). Thus, we first survey the existing privacy filters in a larger domain of videography and then focus on the privacy filters for airborne-recreational videography.

In general, privacy filters can be fixed or adaptable. *Fixed privacy filters* remove sensitive-regions in images or replace them with a de-identified representation. Examples of fixed privacy filters include masking (blanking) [15, 23] and replacing regions representing people with silhouettes or avatars [24–27]. *Adaptable privacy filters* can be configured depending on privacy and fidelity of the targets. Basically, an adaptable privacy filter can be static or dynamic. *Static adaptable filters* keep their parameters, such as the standard deviation of Gaussian blur, spatially and temporally fixed [28, 29]. Examples of static adaptable privacy filters include pixelation [26], Gaussian blur [28] and cartooning [21]. These filters

protect against a naïve attack, but are prone to parrot [12] and reconstruction attacks. *Dynamic adaptable filters* protect sensitive-regions against parrot and reconstruction attacks by changing spatially and temporally their parameters [20, 30], e.g. scrambling [31] and warping [30]. However, these filters introduce flicker as the parameters are temporally uncorrelated.

To the best of our knowledge, there is no dynamic adaptive privacy filter that is designed to reduce flicker and this is first time that flicker reduction is considered in this thesis. Flicker-reduction approaches, which were developed for video compression [32–37], can be classified as *prior*, *during* or *post* encoding [34, 35]. The *prior* [38] and *during* encoding [33–35] approaches are designed for a specific coder. *Post* encoding approaches are instead generic and measure the spatio-temporal correlation between frames [32, 36, 37]. However, this correlation is broken by privacy filters such as scrambling [20] and warping [30]. Therefore an alternative approach for minimising flicker is required for a dynamic adaptive privacy filter.

In particular, privacy filters for aerial videography need to face additional challenges caused by the ego-motion of the camera, changing illumination conditions, and variable sensitive-region orientation and resolution. Recent privacy filters for airborne cameras are based on geo-fencing, Generic Data Encryption (GDE) [39] and Unmanned Aircraft Systems-Visual Privacy Guard (UAS-VPG) [4]. In *geo-fencing*, an MAV avoids to fly over a private property whose co-ordinates could be embedded in the MAV's navigation software (e.g. the community-generated database NoFlyZone [14]) in order to not violate an individual's privacy. Alternatively, in GDE, the MAV first sends *encrypted data* to a privacy server that blanks, blurs or mosaics sensitive-regions and then transfers to an end user. Contrary to server-based filtering, UAS-VPG [4] is aimed for on-board implementation and focusses on face detection in order to Gaussian blur. However, it does not investigate the required Gaussian blur. In fact, faces from airborne cameras can be captured from various angles and distances, thus resulting in a high variation of face orientations and resolutions. If these faces are protected through a fixed Gaussian blur, it can severely degrade the fidelity of the images. Moreover, Gaussian blur as used by both GDE and UAS-VPG is prone to parrot [12] and reconstruction attacks.

In this thesis, we present a privacy-preserving framework for on-board adaptive protection of main-identifiers captured from airborne cameras. Particularly, the two main objectives of this framework are: (i) exploration of Privacy Design Space (PDS) for a given main-

identifier and (ii) its robust protection with minimal spatio-temporal distortion. Although the study and the findings of this thesis are equally applicable to the other main-identifiers, we only consider face of a bystander as it is the most important and well-studied identity source [10, 12, 13, 21, 40–42].

In the first part, the framework employs the existing adaptive privacy filters and explores the PDS for a face. The PDS means a region of the 3D world in which a face is recognisable. We define a mechanism that allows us to automatically configure an adaptive privacy filter. The mechanism uses the resolution of the detected face to determine when it is inherently protected. We use the auxiliary data from the on-board navigation sensors (Global Positioning System (GPS) and Inertial Measurement Unit (IMU)) to determine when a face is not inherently protected and then apply an adaptive privacy filter that distorts a sensitive-region depending upon its captured resolution.

In the second part, the framework is updated with a novel privacy filter called Adaptive Hoping Gaussian Mixture Model (AHGMM) that improves the trade-off between privacy, fidelity and temporal smoothness. In general, the proposed filter distorts a face with secret parameters to be robust to naïve, parrot and reconstruction attacks. The distortion is minimal and adaptive to the resolution of the captured face: we select the smallest Gaussian kernel that reduces the face resolution below a certain threshold. The selected threshold protects the face against the naïve attack as well as maintains its resolution at a specified level. To prevent other attacks, we then insert supplementary Gaussian kernels in the selected Gaussian kernel and hop their parameters locally using a Pseudo-random Number Generator (PRNG) so their estimation is difficult from the filtered face image. Specifically for the videography, hopping of the Gaussian kernels generate flicker just like state-of-the-art scrambling and warping filters. In order to minimise flicker, hopping Gaussian kernels are temporally concatenated with a motion-dependent temporal smoothing filter. In particular, depending on the resolution of the captured face, the parameters of an AHGMM filter are adjusted according to the target spatial distortion and are then temporally averaged with decaying weights to minimise flicker. This minimises spatio-temporal distortions and is robust against naïve, parrot and reconstruction attacks.

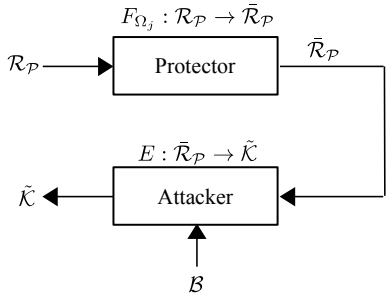


Figure 1.2: Attack model of a privacy-preserving system.

1.2 Problem Formulation

In this section, we discuss the privacy attack model for a protected face and the possible attacks on it. Moreover, we state the performance measures for identity and spatio-temporal distortion for the protected face. Finally, we define the problem of the thesis work.

1.2.1 Privacy Attack Model

Let $\mathcal{D} = \{\mathcal{R}_k\}_{k=1}^{K}$ be a set with face data of K subjects, where k represents the identity information (labels). Let each subject k be represented by at most Y images, i.e. $\mathcal{R}_k = \{R_i | i \leq Y\}$. Let $\mathcal{R}_\mathcal{G}, \mathcal{R}_\mathcal{P} \subset \mathcal{D}$ be the face gallery and face probe data sets, respectively, where $\mathcal{R}_\mathcal{G} \cap \mathcal{R}_\mathcal{P} = \emptyset$ and usually $|\mathcal{R}_\mathcal{G}| > |\mathcal{R}_\mathcal{P}|$, and $|.|$ is the cardinality of a set.
Let $F_{\Omega_j} : \mathcal{R}_\mathcal{P} \rightarrow \bar{\mathcal{R}}_\mathcal{P}$ represent a privacy filter of selected parameter Ω_j. The privacy filter F_{Ω_j} distorts the features of $\mathcal{R}_\mathcal{P}$ to produce a protected probe data set $\bar{\mathcal{R}}_\mathcal{P}$ such that the probability P of predicting its correct labels $\mathcal{K} = \{k_l\}_{l=1}^{|\mathcal{R}_\mathcal{P}|}$ decreases, where l represents the image number of $\mathcal{R}_\mathcal{P}$ and as a result introduce distortion D in each protected face. Let $E : \bar{\mathcal{R}}_\mathcal{P} \rightarrow \tilde{\mathcal{K}} = \{\tilde{k}_l\}_{l=1}^{|\bar{\mathcal{R}}_\mathcal{P}|}$ indicate an attacker whose aim is to correctly predict labels $\tilde{\mathcal{K}}$ of $\bar{\mathcal{R}}_\mathcal{P}$ (see Figure 1.2). In this work, we assume that an attacker has access to $\mathcal{B} \in \{\mathcal{R}_\mathcal{G}, \bar{\mathcal{R}}_\mathcal{G}, \hat{\mathcal{R}}_\mathcal{G}\}$, where $\bar{\mathcal{R}}_\mathcal{G}$ is the filtered gallery data set and $\hat{\mathcal{R}}_\mathcal{G}$ is the filtered and reconstructed gallery data set.
An attacker could modify $\bar{\mathcal{R}}_\mathcal{P}$, $\mathcal{R}_\mathcal{G}$ or both to correctly predict $\tilde{\mathcal{K}}$ of $\bar{\mathcal{R}}_\mathcal{P}$ (see Figure 1.3). In a traditional naïve attack (here referred to as naïve-T attack), a privacy filter is applied on $\mathcal{R}_\mathcal{P}$ to generate a protected probe data set $\bar{\mathcal{R}}_\mathcal{P}$, while the unaltered $\mathcal{R}_\mathcal{G}$ is used for training [12]. A traditional parrot attack (here referred to as parrot-T attack), learns the privacy filter type and its parameters Ω_j (e.g. Gaussian blur of certain standard deviation used to

Figure 1.3: Privacy Attacks to predict label $\tilde{k}$: (a) Naïve-T, (b) Parrot-T and (c) Naïve-SR and (d) Parrot-SR.

generate $\bar{\mathcal{R}}_{\mathcal{P}}$). Then, the learned filter is applied on $\mathcal{R}_{\mathcal{G}}$ to generate a privacy protected gallery data set $\bar{\mathcal{R}}_{\mathcal{G}}$. Finally, $\bar{\mathcal{R}}_{\mathcal{G}}$ and $\bar{\mathcal{R}}_{\mathcal{P}}$ are used for training and testing, respectively [12]. In a reconstruction attack, the discriminating features of $\bar{\mathcal{R}}_{\mathcal{P}}$ are first restored (e.g. using an inverse filter or a super-resolution algorithm) to generate a reconstructed probe data set $\hat{\mathcal{R}}_{\mathcal{P}}$ and then compared against $\mathcal{R}_{\mathcal{G}}$ or a reconstructed gallery data set $\hat{\mathcal{R}}_{\mathcal{G}}$. An inverse filter first estimates the parameters of a privacy filter using $\bar{\mathcal{R}}_{\mathcal{P}}$ and then performs an inverse operation to reconstruct the original faces [18]. Similarly, a super-resolution algorithm first learns embeddings between the high-resolution and their corresponding low-resolution faces and then reconstructs the high-resolution faces for $\bar{\mathcal{R}}_{\mathcal{P}}$ [22]. When such super-resolved faces are compared against $\mathcal{R}_{\mathcal{G}}$ and $\hat{\mathcal{R}}_{\mathcal{G}}$, the attack is referred to as Naïve-SR and Parrot-SR, respectively.

1.2.2 Distortion and Identity Level

A privacy filter F_{Ω_j} generates a spatial as well as a temporal distortion. Let the *spatial distortion level* generated by F_{Ω_j} be measured using the Peak Signal to Noise Ratio (PSNR)

which is given by

$$PSNR = 20\log_{10}\left(\frac{R_{max}}{\sqrt{\Delta R}}\right), \tag{1.1}$$

where R_{max} is the maximum possible pixel intensity and ΔR is the mean square error between the pixel intensities of an unprotected face $R_t \in \mathcal{R}_{\mathcal{P}}$ captured at time t and a protected face $\bar{R}_t \in \bar{\mathcal{R}}_{\mathcal{P}}$ given as

$$\Delta R = \frac{1}{|\mathcal{R}_{\mathcal{P}}|WH}\sum_{r=1}^{|\mathcal{R}_{\mathcal{P}}|}\sum_{w=1}^{W}\sum_{h=1}^{H}|R_t(w,h) - \bar{R}_t(w,h)|_r^2, \tag{1.2}$$

where W and H are width and height of R_t, respectively.

Let the *temporal distortion level* generated by F_{Ω_j} be measured by the maximum of absolute difference ξ of pixel intensities [43] given by

$$\xi = \sum_{w=1}^{W}\sum_{h=1}^{H}\xi(w,h), \tag{1.3}$$

where

$$\xi(w,h) = max\Big(0, |\bar{R}_t(w,h) - \bar{R}_{t-1}(w,h)| - |R_t(w,h) - R_{t-1}(w,h)|\Big). \tag{1.4}$$

where $R_t(w,h)$ ($\bar{R}_t(w,h)$) and $R_{t-1}(w,h)$ ($\bar{R}_{t-1}(w,h)$) are the unprotected (filtered) pixel intensity values from the current and previous frame, respectively.

As there are face recognisers with different recognition capability that an attacker can exploit, it is therefore required to validate F_{Ω_j} using the state-of-the-art face recogniser for the generated spatio-temporal distortion. Let the *identity level* of a bystander be an accuracy $\eta \in \{\eta_v, \eta_i\}$ of a face recogniser, which is calculated depending upon the type of the system, e.g. Equal Error Rate (EER) accuracy η_v for a verification system or cumulative Rank-n accuracy η_i for an identification system. In case of a verification system, η_v is given as

$$\eta_v = \frac{TP + TN}{|\mathcal{R}_{\mathcal{P}}|}, \tag{1.5}$$

where TP and TN are true positives and true negatives, respectively. In case of an identification system, η_i is defined as

$$\eta_i = \sum_{n=1}^{N}\left(\frac{1}{KT}\sum_{k=1}^{K}\sum_{t=1}^{T}x_{kt}\right)_n, \tag{1.6}$$

where N is the identification rank, K is the number of subjects, T is the number of frames in a video, and

$$x_{kt} = \begin{cases} 1 & \text{if } k = \tilde{k} \\ 0 & \text{otherwise.} \end{cases} \tag{1.7}$$

where k and $\tilde{k}$ are the true and predicted labels, respectively.
Let us consider as baseline for comparison a random classifier's accuracy $\epsilon \in \{\epsilon_v, \epsilon_i\}$. For face verification, the random classifier's accuracy $\epsilon_v = 0.5$, while for face identification, a random classifier's accuracy $\epsilon_i = 1/K$. When the face recogniser achieves a performance similar to that of a random classifier, the identity level is considered lowest. If the recogniser performance improves, the identity level increases. Thus, the goal of a privacy filter is to make it harder/impossible for a face recogniser to correctly verify or identify a face or in other words making its accuracy similar to a random classifier.

1.2.3 Problem Statement

The core objective of the thesis work is first to determine whether a captured main-identifier is inherently protected or not. If not, developing a privacy filter with the following two competing targets: The first target is that a face should be robustly protected against attacks, i.e. $\eta \rightarrow \epsilon$ for a protected face under the naïve, parrot and reconstruction attacks. The second concurrent and competing target is that the face should be protected with a minimal spatio-temporal distortion, i.e. high $PSNR$ and low ξ, without affecting the validity of $\eta \rightarrow \epsilon$.

1.3 Research Questions

The thesis addresses the following key scientific questions to protect privacy of the by-standers in recreational videography:

- What is privacy design space for adaptive filtering?
  A camera equipped MAV can manoeuvre in a large space, e.g. up to 150 m in Austria and as a result, could capture a main-identifier from different heights and pitch angles. It could happen that the sensitive-region (i.e. captured main-identifier) is inherently protected due to low resolution or high pitch angle and does not require any privacy filtering. Thus, the MAV should automatically determine whether a sensitive-region

is protected or not. If not, it should adapt the parameters of a privacy filter to the resolution of the sensitive-region in order to maintain high fidelity.

- How to robustly protect a sensitive-region without relying on sophisticated detectors?
  A sophisticated detector means any computer vision algorithm other than the object detection algorithm, e.g. pose, age, race, gender detectors are termed as sophisticated detectors when used along with a face detector. An adaptive privacy filter, e.g. Gaussian blur is prone to the parrot and the reconstruction attacks, while, in order to protect against such attacks, a non-adaptive privacy filter, e.g. k-Same-Select [44], k-Same-M [40], and Gender Age Race Protector (GARP) [41] require all the k-sensitive-regions to be at the same pose. Finding the exact pose of a main-identifier from its given sensitive-region in airborne videography could be challenging due to illumination conditions and variable resolutions. Thus, it is required to develop a robust privacy filter that does not depend on any sophisticated detector.
- How to minimise spatio-temporal distortion in privacy-preserving airborne videography?
  An MAV can capture a main-identifier with variable resolution and pitch angle in a video. Applying an adaptive privacy filter with constant parameters could reduce fidelity as the pitch angle or height of the camera increases or simply the resolution of the sensitive-region decreases. In addition, the switching of the parameters for different frames of a video generates flicker, e.g. as in scrambling [20] and warping [30]. This generates unpleasant effect and can also divert a viewer's attention. Thus, it is required to adapt the parameters of a filter depending on the resolution of the sensitive-region without creating any temporal distortion. Moreover, the parameter adaptation should not affect robustness against different attacks.

1.4 Contributions

The key contributions of the thesis are:

- *Privacy design space for adaptive filtering:* Design and development of a mechanism to determine whether a sensitive-region is inherently protected or not and if not, adaptively selecting the filter parameters to minimise distortion. The mechanism is based on the pixel density (i.e. number of pixels per unit distance) of a sensitive-region and a pre-defined threshold. The pixel density is estimated using the height, tilt angle

and image-sensor dimensions of the camera, while the threshold is experimentally calculated for a state-of-the-art recognition algorithm, e.g. a face recogniser for a captured face. When the sensitive-region is recognisable in the image/video, the estimated pixel density also assists in optimally configuring an adaptive privacy filter to protect the sensitive-region with a minimal distortion. The salient feature of the optimal configuration is that it provides privacy equivalent to blanking out (only against a naïve attack) but much higher fidelity compared to fixed filtering or blanking out. *Deliverables:*

O. Sarwar, B. Rinner, and A. Cavallaro. Design space exploration for adaptive privacy protection in airborne images. In Proc. IEEE Advanced Video and Signal-based Surveillance (AVSS), pages 159-165. Colorado Springs, USA, August 2016.

- *Adaptive Hopping Gaussian Mixture Model (AHGMM):* Design and development of a novel privacy filter that robustly protects a sensitive-region against brute force, naïve, parrot and reconstruction attacks, and at the same time minimises spatio-temporal distortion. The salient features of AHGMM are: (1) It does not rely on any sophisticated detector for robust protection in contrast to non-adaptive privacy filters: k-Same-Select [44], k-Same-M [40], k-Same-Furthest [45], k-Same-Net [46] and GARP [41]. (2) Like adaptive privacy filters (Space Variant Gaussian Blur (SVGB) [47], Optimal Distortion-Based Visual Protector (ODBVP) [13], Cartooning [21]), the AHGMM can adaptively be configured for a better trade-off between fidelity and privacy protection. However, unlike these adaptive privacy filters, protection of AHGMM is robust against parrot and reconstruction attacks. (3) Finally, compared to dynamic adaptive privacy filters (scrambling [20, 48–51] and warping [30]), the AHGMM minimises flicker by temporally correlating its parameters without compromising on robustness. *Deliverables:*

  O. Sarwar, A. Cavallaro, and B. Rinner. Temporally smooth privacy protected airborne videos. In Proc. IEEE International Conference on Intelligent Robots (IROS), Madrid, Spain, October 2018.

  O. Sarwar, B. Rinner, and A. Cavallaro. Concealing the identity of faces in oblique images with adaptive hopping Gaussian mixtures, International Journal of Computer Vision (IJCV), Springer (planned, preprint at http://arxiv.org/abs/1810.12435).

- *Face data sets:* Generation of a synthetic face image data set (see Appendix A) and collection of a face video data set (see Section 5.2.1). To the best of our knowledge,

there is no publically available face image data set with a large population size or a face video data set even with a small population size captured from an MAV. In this thesis, we synthetically generate a face image data set of 480,000 images belonging to 4281 subjects which emulates faces as captured from an MAV at different heights and pitch angles. In addition, we collect a small face video data set of 11 subjects from two different heights (4 m and 7 m) with a pitch angle variation of 20° - 78°, which shows faces of moving subjects as captured from an MAV. *Deliverables:*
Airborne face image data set (Appendix A) and airborne face video data set (Section 5.2.1).

1.5 Thesis Outline

The rest of the thesis is organised as:
Chapter 2 first describes the privacy definition used in the visual data and then critically reviews the state-of-the-art privacy filters, including both airborne and CCTV cameras. Specifically for the human faces, it states the recognition requirements for humans as well as machine algorithms. In the end, it summaries the state-of-the-art privacy filters by highlighting the research gaps.
Chapter 3 presents the concept of PDS exploration. Moreover, when operating inside the explored PSD, it describes how an adaptive filter could be optimally configured to increase fidelity without compromising on privacy. It experimentally supports and validates optimal configuration using a Gaussian blur.
Chapter 4 presents a novel privacy filter based on hopping kernels to improve the trade-off between privacy, fidelity and flicker. In particular, the chapter describes the concept of hopping kernels and how to generate them for a given sensitive-region such that it is robustly protected with a minimum spatial distortion. Moreover, the chapter gives the details of spatio-temporal post-processing aimed for seamless spatial protection and reduced temporal distortion. In the end, it presents the analytical analysis of the computational complexity of the proposed filter.
Chapter 5 discusses the experimental results of the proposed filter for airborne photography and videography by exploiting a synthetic and a real face data sets, respectively. The chapter assumes different knowledge levels of an attacker for the parrot and the reconstruction attacks and thoroughly investigates privacy. For the reconstruction attacks, it uses

an inverse filter and a state-of-the-art super-resolution algorithm. Moreover, it quantifies the corresponding fidelity and flicker. In the end, it presents the trade-off analysis between privacy and fidelity for a photography scenario, and between privacy, fidelity and flicker for a videography scenario.

Chapter 6 finally concludes the thesis by summarising the key contributions, highlighting the limitations and elaborating the future research directions.

Chapter 2

State of the Art

Privacy protection in airborne cameras is a contemporary research area with only a few frameworks mostly based on access control. However, privacy protection methods developed over the last two decades for ground cameras, i.e. CCTV could prove useful for airborne cameras. A direct application of these methods on airborne cameras can be insufficient as airborne cameras introduce additional privacy challenges of mobility and viewing angles. Therefore, this chapter discusses privacy-preserving filters proposed for both CCTV and airborne cameras, and then highlights their limitations.

First, this chapter discusses the definition of privacy and utility used in the literature of visual data and describes their important aspects for recreational videography. Second, it formally defines a privacy filter and critically reviews privacy filters proposed for both CCTV and airborne cameras, highlighting whether these filters preserve privacy and maintain utility according to their stated definitions or not. At the end of the review process, it elaborates the limitations of the existing state-of-the-art to protect privacy in airborne cameras and states the differences of the proposed work. Specifically, considering face as a main-identifier, it discusses the face recognition requirements for both humans and machine algorithms. Finally, the chapter concludes by presenting a summary.

2.1 Definitions

2.1.1 K-anonymity

A k-anonymity algorithm takes the k-attributes (e.g. faces) belonging to k-individuals of a data set and replaces them with a single attribute, ensuring that the attribute is not recognised with a probability greater than $1/k$. Increasing the value of k improves privacy protection, but reduces the discriminability of the data that may limit its usefulness. Examples of k-anonymity algorithms for face data sets are k-Same [12], k-Same-Select [44], k-Same-M [40]. In contrast to k-anonymity, anonymisation, in general, can also be achieved through ad-hoc algorithms, e.g. blurring [28], pixelation [26] and cartooning [21], which distort the faces to impede their recognition. Through the parameters of these algorithms, different levels of anonymisation can be ensured.

2.1.2 Differential Privacy

Both k-anonymity and ad-hoc algorithms ensure protection within a given data set captured under a certain context. However, when two such anonymised data sets with a minimum difference are compared, individuals can be recognised with high probability [52]. To protect such identity leakage, differential privacy algorithms [53, 54] are proposed which reduce the difference between the two anonymised data sets, thus providing differentially protected data sets. In this thesis, we focus on a single data set of bystanders' faces; therefore, we do not consider differential privacy protection.

2.1.3 Privacy Loss

There are different definitions of privacy loss as it is a subjective issue [10, 15, 55]. In this thesis, privacy loss L is defined as a probabilistic variable [9, 10] given as

$$L = I.S, \tag{2.1}$$

where $I \in [0, 1]$ is the probability of identifying a bystander using any possible source (e.g. face, vehicle licence plate, age, race, location and time) and $S \in [0, 1]$ represents the sensitivity index of an activity (i.e. an amount of personal information in the activity as perceived by the bystander and 1 being highly sensitive). This activity could be any personal habit, e.g. smoking, playing any sport or presence at some location.

Identity is an objective entity in L and can be estimated using main-identifiers (i.e. face, ear, fingerprint, vehicle licence plate), quasi-identifiers (i.e. age, race, gender, hair-style, type and colour of clothes) and context (i.e. position and time) [10]. If I_m, I_q and I_c are the identity probabilities estimated through the main-identifiers, quasi-identifies and context, respectively, then I is given by

$$I = \nu_m I_m + \nu_q I_q + \nu_c I_c, \tag{2.2}$$

where $0 \leq \nu_m, \nu_q, \nu_c \leq 1$ are weighting factors such that $\nu_m + \nu_q + \nu_c = 1$. An attacker can adaptively select ν_m, ν_q, ν_c to maximise I depending on the number of bystanders in a public place and their gender, age, race, colour and types of clothes.

In contrast to identity, the sensitivity index is a subjective entity [15], which depends on the bystander, his/her cultural background, religious values [55] and age [9]. Moreover, the sensitivity index of an information also depends on its context [55]. For example, an information might not be sensitive in one context (home), but highly sensitive in another context (work). A bystander may also grade the same activity differently during his/her teenage, young age and old age. Let $A_{t,\Delta t}$ be an activity starting from time t for Δt duration and possesses certain personal information. The sensitivity index of $A_{t,\Delta t}$ is given by

$$S = O(A_{t,\Delta t}), \tag{2.3}$$

where O is a subjective operator which grades $A_{t,\Delta t}$ depending upon the amount of personal information.

In order to minimise L, either I or S has to be reduced. In this thesis, we only reduce I due to its objective nature compared to subjective S. In particular, we reduce I_m as it carries more importance in comparison to I_q and I_c, especially when the number of bystanders increases in a public place. We measure and validate I_m using Eq. 1.5 or Eq. 1.6 considering only faces as main-identifier.

2.1.4 Utility

When a privacy filter is applied to protect the main-identifiers, quasi-identifiers or context, it introduces a distortion which may limit the use of the protected images and videos. As a result, there could be a trade-off between privacy loss and utility of the images and videos. In fact, utility refers to the usefulness of the privacy-protected images and videos, and can

be stated depending upon the application. For example, in an application that requires only detection or tracking of human faces and not their recognition, utility can be defined as the accuracy of a detection or tracking algorithm [42]. Similarly, in an application (e.g. recreational videography) where the visual quality of the protected image is more important, utility can be defined as fidelity [21, 56]. Although the quasi-identifiers can help in inferring the identity of bystanders, they can also represent utility in some applications, especially where the population size is large and accurate estimation of the quasi-identifiers does not contribute too much in identifying bystanders. For example, it may be required to differentiate the faces based on age, race or gender in a large face data set while still thwarting their recognition [41].
In this thesis, we consider the application of recreational videography using an MAV; therefore utility is defined as the fidelity (measured using Eq. 1.1) of the protected videos as well as their flicker level (i.e. temporal distortion level measured using Eq. 1.3).

2.1.5 Privacy Filter

A privacy filter is either an arrangement of optics of a camera or a computer vision algorithm that modifies the appearance of critical-areas of an image such that the underlying identity or activity is difficult to recognise by machine algorithms or humans. The critical-areas can be either sensitive-regions that represent the captured main-identifiers, quasi-identifiers or background that shows the captured context. A privacy filter can be applied either during the image capturing stage (pre-processing) or after capturing the image (post-processing). In addition, a privacy filter can be either reversible or irreversible [9, 57]. A reversible privacy filter can recover back the original critical-areas without any loss, while an irreversible privacy filter cannot fully recover back the original critical-areas. Moreover, a privacy filter can be either adaptive or non-adaptive [13, 21]. An adaptive privacy filter can control the distortion strength while protecting a critical-area, while a non-adaptive privacy filter cannot control the distortion strength. Finally, a privacy filter can be manual, semi-autonomous or autonomous: in a manual privacy filter, a bystander is himself/herself responsible for his/her privacy protection without camera involvement; in a semi-autonomous privacy filter, the bystander interacts with the camera and directs the camera to protect privacy; in an autonomous privacy filter, the camera itself finds and protects the bystander. Table 2.1 presents a summary of the state-of-the-art privacy filters.

Table 2.1: State-of-the-art for privacy protection in visual data. Implementation stage means position in the image capturing pipeline where a filter operates and it can be pre-processing (before capturing an image) and post-processing (after capturing the image). Target area indicates whether the filter protects sensitive-regions, context or both. Protection type represents properties of the filter: adaptive means controllable distortion, while irreversible means distortion cannot be removed. Autonomy indicates degree of interaction between a camera and a bystander, which can be manual (the bystander covert himself/herself from being captured), semi-autonomous (the bystander directs the camera to protect him/her) and autonomous (the camera itself first finds and then protects the bystander).

References	**Implementation stage**		**Target area**		**Protection type**		**Autonomy**		
	Pre-processing	Post-processing	Sensitive-region	Context	Adaptive	Irreversible	Manual	Semi-autonomous	Autonomous
[58, 59]	✓		✓	✓		✓	✓		
[60, 61]	✓		✓			✓	✓		
[39, 62, 63]	✓		✓			✓		✓	
[64–67]	✓		✓			✓			✓
[14, 68–70]	✓			✓		✓			✓
[71]	✓			✓		✓		✓	
[13, 20, 30, 31, 48–51, 72–82]		✓	✓		✓				✓
[19, 83–86]		✓	✓						✓
[4, 13, 17, 26, 28, 39, 47, 79, 87]		✓	✓		✓	✓			✓
[9, 11, 12, 23–27, 40, 41, 44–46, 79, 88–97]		✓	✓			✓			✓
[21]		✓	✓	✓	✓				✓
[98]		✓	✓	✓					✓
[15]		✓	✓	✓					✓
[16, 99]		✓		✓		✓			✓

2.2 Pre-processing Privacy Filters

Pre-processing privacy filters prevent a camera to capture main-identifiers, quasi-identifiers or even context during the image acquisition. These privacy filters can be divided into sensitive-region and context-oriented filters. Both state-of-the-art sensitive-region and context-oriented filters are irreversible and non-adaptive.

2.2.1 Sensitive-Region Privacy Filters

Sensitive-region pre-processing privacy filters only protect the visual details of the main-identifiers or quasi-identifiers during image acquisition phase. This protection can be enabled by interacting with a camera through either its software or hardware. Based on the type of interaction between the bystander and the camera, sensitive-region pre-processing privacy filters can further be divided into manual, semi-autonomous and autonomous categories.

A manual sensitive-region privacy filter usually protects main-identifiers or quasi-identifiers by forcefully saturating the camera sensor; therefore the responsibility of privacy protection relies on the bystander. For example, in flash photography, Eagle Eye [58] uses a light detector to locate a camera and then bursts intense light on it. Similarly, BlindSpot [59] blocks the cameras, but it does not rely on flash detection. Rather, it actively searches for retro-reflective Charge Coupled Device (CCD) or Complementary Metal Oxide Semiconductor (CMOS) camera lenses using Infra-red (IR) illuminators. Both Eagle Eye and BlindSpot non-adaptively distort almost the complete image which significantly reduces the fidelity. On the contrary, Privacy Visor [60] is specifically designed for face protection and only distorts the sensitive-region (i.e. captured face). It consists of a pair of glasses equipped with Near Infra-red (NIR) Light Emitting Diodes (LEDs). These LEDs continuously emit invisible light which saturates the image sensor that tries to capture the face. In contrast to Privacy Visor, LiShield [61] uses a specialised pulsating LEDs to protect the face from attackers, but it allows the authenticated users to capture it by sharing the pulsating pattern through wireless communication. Due to the pulsating nature of the LEDs, LiShield only corrupts different sub-regions of the captured face. However, faces protected by both Privacy Visor and LiShield still remain recognisable for humans.

The motivation of semi-autonomous sensitive-region privacy filters is to disable the cameras or notify about photography prohibition through wireless communications, e.g. Blue-

tooth [62], Wi-Fi [39] and IR [63]. Both Bluetooth [62, 100] and Wi-Fi [39] based approaches use standard communication protocols and therefore require special software installation on the camera to control it. Thus, these approaches completely rely on the integrity of the camera owner. In contrast, an IR-based approach [63] controls the camera through a dedicated IR-receiver that is fabricated within the camera. In this approach, a main-identifier encodes its privacy policy in the IR signal which is transmitted and finally decoded by the camera's IR-receiver to switch it on or off. Due to communication constraints, these privacy filters are only suitable for a short range.

An autonomous sensitive-region privacy filter does not require any involvement of the bystander, and autonomously first detects and then protects his/her main-identifiers at the image sensor level. For example, the "anonymous camera" is fully autonomous and consists of an IR sensor and a CCD sensor along with a Spatial Light Modulator (SLM) placed in front of it [64]. The camera detects the face using the IR sensor and then projects it on the SLM, which optically blanks out the face for the image captured by the CCD sensor (see Figure 2.1). Recently, a prototype of pre-processing k-anonymity is developed that optically k-anonymises a face [65]. However, it first requires to find the exact pose of the face and then requires the $k-1$ faces at the same pose.

In contrast to the anonymous camera, the courteous-wearable camera developed by Microsoft fabricates a Far Infra-red (FIR) image sensor with an RGB image sensor to autonomously control the RGB image sensor without using any SLM [66]. Initially, the RGB sensor is kept off and only the FIR image sensor captures and analyses the environment using computer vision algorithms. If it does not find any privacy violation, the FIR image sensor switches on the RGB sensor; otherwise, the RGB image sensor remains off. The motivation of the FIR image sensor to control the RGB sensor is its low quality which is less privacy invading but sufficient to sense the environment. In the same stream, the PrivacEye wearable camera uses two RGB sensors: one controlling sensor and second recording sensor [67]. The controlling sensor is placed close to the human eye which analyses the eye movements to control the shutter of the recording sensor. It is argued that eye movements are related to private clues and therefore could be used to preserve the privacy. However, such protection mechanisms only allow capturing a video at irregular intervals.

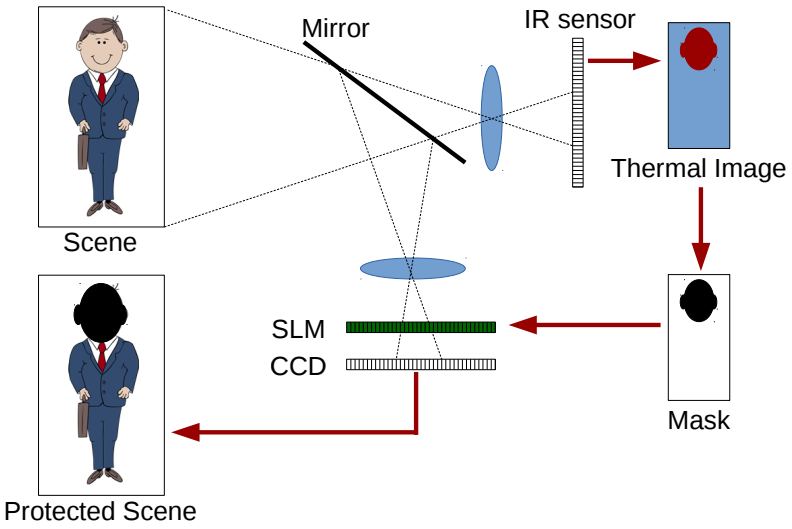


Figure 2.1: An autonomous sensitive-region pre-processing privacy protection filter. A sensitive-region (i.e. face) region is first detected using an IR sensor which is then projected on the SLM sensor, placed in front of the CCD sensor. The SLM sensor blocks the light rays corresponding to the detected face region from falling on the CCD sensor. Thus, the face region is blanked out on the CCD sensor. This figure is an adaptation from [64].

2.2.2 Context-Oriented Privacy Filters

Context-oriented pre-processing privacy filters protect the background (i.e. captured context) of the images and videos. The background contains the location and time information which could be exploited to infer the identity of a bystander, even when his/her main-identifiers or quasi-identifiers are unrecognisable [9, 47]. These filters are usually aimed to protect the private spaces, e.g. backyard of a home or view of a room as seen through its glass window. Particularity for airborne cameras, the context-oriented pre-processing privacy filters are based on the idea of restricting the MAVs from accessing the private spaces either indefinitely or for a specified time. These filters are irreversible and can be categorised as semi-autonomous and autonomous privacy filters.

Autonomous context-oriented privacy filters automatically detect the private spaces and protect them without any real-time involvement from the owner of the context, i.e. no real-time request to protect his/her context. For example, NoFlyZone [14] encloses a database of restricted areas in the navigation software of an MAV and avoids these restricted areas while

planning the MAV's route. A practical implementation of this is a B4UFLY app developed by Federal Aviation Administration (FAA) to assist drone pilots. However, such context-oriented pre-processing privacy filters are insufficient to protect privacy, especially when hovering near the residential areas. For example, consider an MAV flying in an allowed region (public park) with adjacent NoFlyZone area (home), but pointing its lens in the NoFlyZone property.

Contrary to property restrictions, it is also suggested to put restrictions on just altitudes irrespective of the horizontal location to prevent high-resolution images [70]. In particular, it is advocated to extend landowner's airspace right in the United States up to 350 ft (currently 83 ft), just 50 ft below the maximum allowed altitude (400 ft) for the commercial MAVs [70]. The rest of 50 ft spatial window (350-400 ft) is dedicated for the transportation of commercial MAVs. However, restrictions on just height can leak privacy, e.g. using a high-resolution camera/optics.

Semi-autonomous context-oriented privacy filters protect the context by incorporating context-owner-defined privacy policies that are directly communicated to the MAV. For example, Mind-your-(R, ϕ) advertises privacy policies associated with a private property through a WAP-beacon over a Wi-Fi link [71] (See Figure 2.2). An MAV approaching a private property listens to the beacon signal and determines whether it is allowed to use the in-coming private space or not. If allowed, it decodes additional constraints: height, time, noise level and sensors. Finally, based on these constraints, it updates its route and heads towards its destination. Like an autonomous context-oriented privacy filter, this approach is also not suitable in case of oblique optics.

2.3 Post-processing Privacy Filters

Post-processing privacy filters protect critical-areas (i.e. captured main-identifiers, quasi-identifiers and context) after image acquisition. All post-processing privacy filters are autonomous, i.e. they do not require any interaction from bystanders. These privacy filters, like pre-processing privacy filters, can be divided into sensitive-region and context-oriented categories.

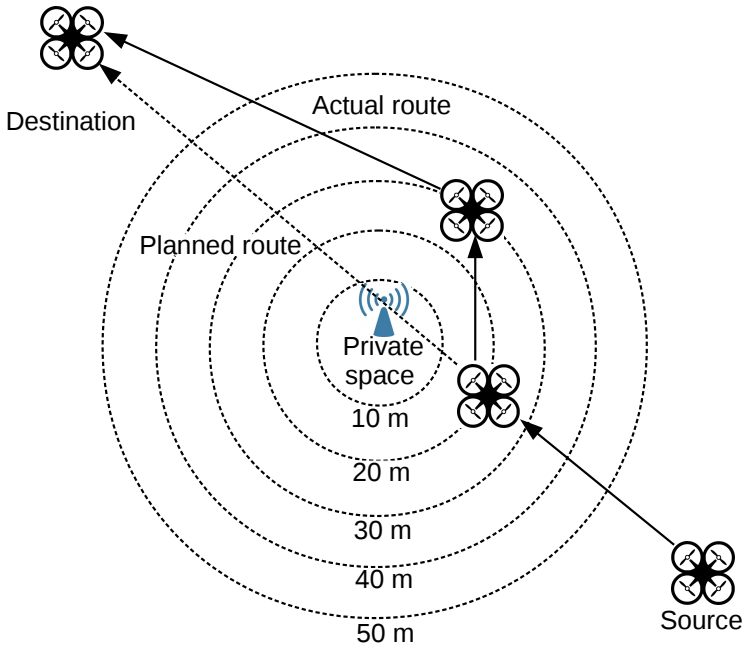


Figure 2.2: A real time property-restricted privacy protection approach. An MAV is required to fly from a source to a destination location, but a private property is located in the direct shortest path. When the MAV approaches the private property, it listens a WAP-beacon signal broadcasted by the private property. After decoding the signal, MAV finds that it is not allowed to traverse the private property at the moment and therefore reroutes its path. This figure is an adaptation from [71].

2.3.1 Sensitive-Region Privacy Filters

Sensitive-region post-processing privacy filters protect the sensitive-regions (i.e. captured main-identifiers or quasi-identifiers) in the images and videos using computer vision algorithms. These privacy filters can be further divided into reversible and irreversible filters.

Reversible privacy filters: These filters use a private key to protect the sensitive-regions, which is later used to recover back the original sensitive-regions if required. These filters can be either adaptive or non-adaptive.
Adaptive reversible filters protect the sensitive-region with a controllable distortion which can be fully removed with the secret key. Examples of adaptive reversible filters are scrambling [20, 31, 48, 72–74], warping [30] and morphing [76].

(a) (b)

Figure 2.3: Privacy protection through invertible cryptographic obscuration. Although the visual distortion can be fully removed, it leaves a sensitive-region (e.g. a captured face) incomprehensible when protected.

Scrambling can be used in space, frequency or code-stream domain [9]. The main advantage of the space domain scrambling is its simplicity and independency of the used compression algorithm [74, 75]. However, space domain scrambling increases the bit rate as it changes the data statistics significantly. Contrary to the space domain, a huge amount of work exists for transform domain scrambling. Examples of transform domain scrambling are: permuting the magnitude of the Discrete Cosine Transform (DCT) coefficients [72, 73]; only changing the sign of the DCT coefficients [31]; permuting the magnitude as well as inverting the sign of the coefficients depending upon the frequency contents [48]. However, such DCT-based-scrambling provides a weak protection as fixation of DC-coefficients and changing AC-coefficients could help to recognise a protected face [82, 101, 102]. To provide a strong scrambling effect in H.264/AVC standard, different approaches are suggested. For example: encrypt the sign of non-zero coefficients as well as Intra Predicted Modes (IPM) [49]; modify the IPM of smaller blocks (4×4) and Motion Vector Difference (MDV) of larger blocks (greater than 4×4) [50]; first encrypt the DC coefficients and then permute both DC and AC coefficients [82]. However, it is shown that the transform domain scrambling degrades coding efficiency. In order to avoid degradation of coding efficiency and full decode/encode of the video for privacy protection, it is proposed to perform scrambling in the code-stream for MPEG-4 [20]. However, code stream domain scrambling does not guarantee that it

would be decodable by the standard player.

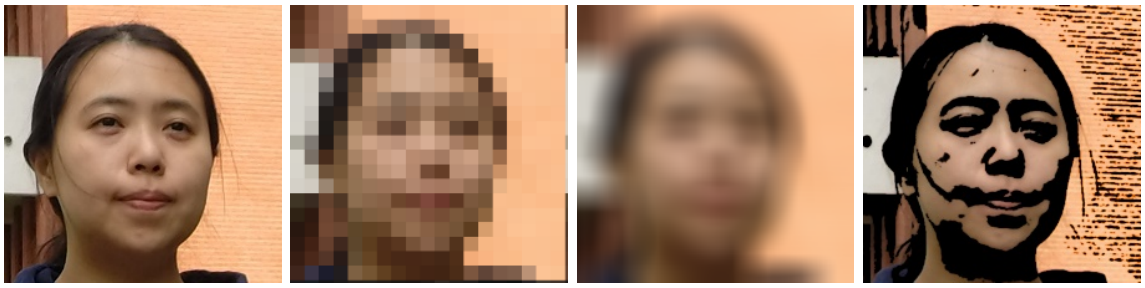
Other adaptive reversible filters like warping [30] and morphing [76] only work in the space domain. Warping shuffles the co-ordinates of a face by exploiting key pixels, while morphing blends an original source face with a target (reference) face by selecting key points in both faces. Recently, in order to preserve facial expressions in morphing, image melding is exploited [78, 103]. The limitation of warping is that it generates flicker while morphing requires both faces at the same pose.

Non-adaptive reversible filters distort the sensitive-regions but cannot control the distortion strength. These filters are usually based on generic encryption (i.e. Data Encryption Standard (DES) and Advanced Encryption Standard (AES)), and chaotic encryption. Examples of privacy filters using these encryption algorithms are: Privacy through Invertible Cryptographic Obscuration (PICO) [19, 83], Privacy using Chaos Cryptography (PCC) [84] and Privacy by Trusted Computing (PTC) [85]. Depending upon the compression format, PICO encrypts the sensitive-region in the spatial domain or frequency domain and also presents a prototype "PrivacyCam" [83]. Similarly, PTC uses a Trusted Platform Module (TPM) microchip to provide hardware enabled privacy protection through AES algorithm [85]. Instead of traditional cryptography, PCC evaluates the chaotic cryptography to conceal a sensitive-region [84]. It is shown that the chaos cryptography is computationally more efficient compared to traditional cryptography as the size of the sensitive-region is increased. However, the main disadvantage of traditional as well as chaotic encryption is that it leaves sensitive-region incomprehensible.

Another non-adaptive reversible filter has recently been developed using false colours which has been evaluated only against a naïve attack [98]. In this filter, the RGB pixel intensities belonging to the sensitive-region are first converted into grey scale which are then mapped to another RGB pixel intensities using custom colour pallets. To make it reversible, the vector difference of the original RGB and mapped-RGB pixel intensities are encoded in the JPEG format.

Irreversible privacy filters: These privacy filters either permanently deform the features of a sensitive-region or replace it with another de-identified sensitive-region. Irreversible privacy filters, like reversible filters, can also be further categorised into adaptive and non-adaptive filters.

Adaptive irreversible privacy filters tend to remove the high frequency contents of a sensitive-



(a) (b) (c) (d)

Figure 2.4: An image protected with irreversible privacy filters: (a) Original (source [104]) (b) pixelation, (c) Gaussian blur and (d) cartooning.

region which are usually exploited for identification. These filters usually operate in the spatial domain and distort the pixel intensities of a sensitive-region which are not fully re-storable. Examples of adaptive irreversible privacy filters are pixelation [26], Gaussian blur [28] and cartooning [21] (see Figure 2.4). Pixelation computes the average of the pixel intensities in the predefined kernel and then replaces all the pixel intensities with their average value. In contrast, Gaussian blur computes individual pixel-intensity using the weighted neighbouring pixel intensities defined by the Gaussian function. Cartooning filter first applies a mean shift algorithm to cluster colour information and then superimposes separately detected edges.

Due to the simplicity of the adaptive irreversible filters, different privacy-preserving frameworks using these filters have recently been proposed for MAVs. These frameworks apply adaptive irreversible privacy filters at the ground station [79], server [39] or on-board [4]. Ground station based frameworks apply these filters after downloading the videos at the end user's device [79]. Such a framework is perceptible to attacks while the videos are being downloaded and therefore could leak privacy. To resist attacks, a server-based privacy-preserving framework is presented in which MAVs send encrypted videos to the server for filtering the sensitive-regions before forwarding them to the end users [39]. The server-based approach could cause latency due to traversing the videos between an MAV, server and end user. To overcome this latency issue, UAS-VPG [4] aims for on-board implementation which focusses on face detection from MAVs in order to apply Gaussian blur. However, it has been shown that Gaussian blur and pixelation are weak privacy filters [12, 40, 105]. That means, if an attacker successfully estimates the filter parameters, original sensitive-regions

(a) k=2 (b) k=4 (c) k=6 (d) k=8 (e) k=10

Figure 2.5: Faces de-identified using k-Same algorithm [12] using different values of k. (a-e): First row represents the original images, while the second row shows de-identified faces. As the k-Same algorithm does not consider pose, gender or facial-expression for the de-identification, it results into poor quality of de-identified faces.

can be recognised with high accuracy. Although such an attack has not been studied for the cartooning filter, some privacy leakage is expected as well.

Non-adaptive irreversible privacy filters either remove a sensitive-region [11, 88–90] or replace it with an abstraction to maintain aesthetic pleasantness or behaviour information [9, 15]. Different abstraction methods have been proposed to replace a human body, e.g. bounding box, avatars, edge, transparency, silhouettes [15, 24–27]. To protect clothing and skin colour, a neural-network is recently exploited to transfer the colour of a style image to the detected human body [97, 106].

Specifically for human faces, face de-identification algorithm k-Same [12] replaces the captured k-faces with their average face to maintain k-anonymity (see Figure 2.5). However, it does not maintain pose, facial expression and gender information. To maintain pose, facial expression and gender information in the protected faces, k-Same-Select [44] and k-Same-M [40] algorithms are proposed, which exploit pose, facial expression and gender detectors. Recently, a number of variants of these algorithms have been developed in order to preserve pose, facial expression, gender, race and age using their respective detectors [41, 91, 92].

To provide recognition accuracy less than 1/k where k is the number of faces in a cluster, k-Same-furthest algorithm [45] is proposed which is basically an extension of k-Same. It

iteratively forms two clusters of k-Same faces and then replaces each captured face in the first cluster with the average of the second cluster and vice versa. Another extension of k-Same is the k-Same-Net algorithm [46] that applies a Generative Neural Network (GNN) on the k-Same algorithm to generate privacy-preserved faces while maintaining the attributes of the unprotected face, e.g. facial expression, age and gender.

Face de-identification can also be achieved by changing the projection bases. For example, arbitrarily modifying the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) projections of a captured face can protect it from face recognition algorithms under the naïve attack [94]. Similarly, the captured face could be de-identified by projecting it on a hypersphere of a radius that is estimated using the gallery faces [95, 96].

Although all the face de-identification algorithms based on k-anonymity are aesthetically pleasant and also exhibit attacks resilience, their suitability highly depends upon sophisticated visual detectors, i.e. pose, gender, race, age or facial-expression whose accuracy depends upon face resolution, face pose and illumination conditions which are often critical challenges in airborne photography and videography.

2.3.2 Context-Oriented Privacy Filters

Context-oriented post-processing privacy filters protect the identity leakage from the background (i.e. captured context) of a video after capturing the video. In particular, these filters aim to distort or remove any location or time related clue from the background of the images and videos that may help in inferring the identity of a bystander. Context-oriented post-processing privacy filters are autonomous and can be divided into reversible and irreversible categories.

Reversible Context-Oriented Privacy Filters

These filters protect the background of the images and videos with an aim to fully recover back if required. A comprehensive design of a privacy protected framework with fully recoverable background is presented by Senior et al. [15]. During the video analysis, different information streams carrying background and foreground objects are generated which are further encrypted for conditional access (see Figure 2.6). To recover back original video including the background, all the information streams are decrypted and foreground objects are placed back in the background image.

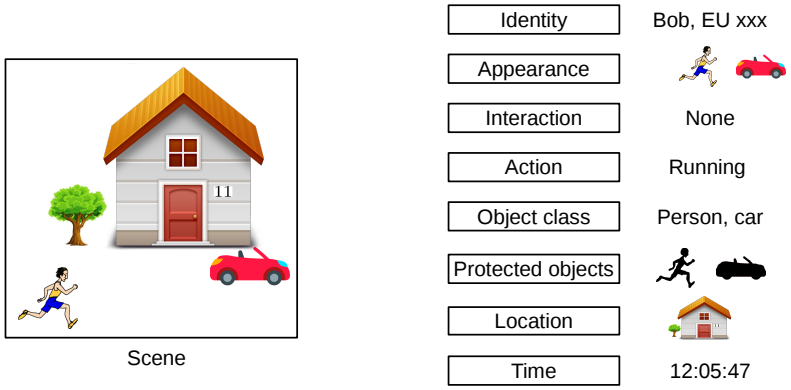


Figure 2.6: Reversible foreground and context protection in videos. The left side shows a video frame which could be rendered based on foreground and contextual objects shown on the right side. All the original foreground and contextual objects are encrypted and could recover the complete left image, if required. This figure is an adaptation from [15].

Irreversible Context-Oriented privacy filters

These filters protect the background by using existing irreversible algorithms e.g. blank out, pixelate or Gaussian blur. In particular, these filters protect any clue of location or time information captured in the background that could assist in identifying bystanders [16]. Through subjective evaluation, it is shown that if viewers do not have acquittance of the context and the captured context i.e. the background is protected, they are unable to determine identities of bystanders. However, protecting only the textual clues of location and time information in the background has a serious drawback as the advanced computer vision algorithms can classify background into common areas, bedrooms, bathrooms, office with high accuracy [99]. To the best of our knowledge, it is the only irreversible context-oriented privacy filter.

However, a lot of work has been carried out to protect spatio-temporal information of bystanders in Location Based Service (LBS) and participatory sensing field which use non-visual data, e.g. WLAN, 3G/LTE, Bluetooth. This work can be broadly divided into location and trajectory protection [107]. Location protection is usually achieved by inserting dummy locations instead of originals [108], spatio-temporal cloaking to achieve k-anonymity [109, 110] and intentionally degrading location/time information [111]. The trajectory protection approaches usually protect the complete route, i.e. origin, intermediate and desti-

nation points by developing dummy trajectories [112], and grouping trajectories to achieve k-anonymity [107, 113–115].

Visual context protection is a great challenge in both fixed and airborne cameras. Applying LBS and participatory sensing approaches to the visual data and at the same time maintaining high fidelity becomes very difficult and challenging task. Thus, it is one of the open research areas to focus.

2.4 Resolution for Face Recognition

Face recognition is about determining the identity information of an individual [116, 117]. Face recognition can be performed either by humans or machine algorithms. Both have their own constraints to accomplish it. Parameters such as camera resolution on the target, camera view angle, occlusion, lighting conditions accumulate uncertainty in fulfilling this recognition task. Among these, we explore minimum resolution required for the face recognition by both humans and machine algorithms.

For a human, considering full human body, British Security Industry Association (BSIA) and European Committee for Standardization (CEN) have stated minimum pixel density and body to viewing screen ratio for face recognition [118, 119]. In contrast, Axis Communication (AC) argues that for digital cameras just horizontal pixel density is sufficient and explicitly states the required horizontal pixel density for the face recognition [120] which is given in Table 2.2.

AC states the horizontal pixel density but not the required vertical pixel density. The reason might be that mostly CCTV cameras are installed at low height with a slight tilt angle. Therefore, the stated pixel density in the horizontal direction also ensures sufficient pixel density in the vertical directions. This can be acceptable in the CCTV environment, but in airborne videos it is not always true. For example, for a directly downward looking camera, we might achieve the stated horizontal pixel density, but with a very low pixel density in the vertical direction. Therefore, for airborne videos, minimum vertical pixel density is also required in addition to the horizontal one.

For a standard stated horizontal pixel density, corresponding vertical pixel density can be estimated through camera dimensions and assuming an angle of about 20° between ceiling

Table 2.2: Face recognition requirements for a human face by a human operator and a machine algorithm (PCA/LDA based face recognizer for accuracy greater than 70%). In this table, a face width of 16 cm has been assumed to determine horizontal number of pixels.

Objective	Human Operator [120]		Machine Algorithm [121]	
	Pixels per face (horizontal)	Pixel Density (px/cm)	Pixels per face	Pixel Density (px/cm)
Identification (good)	40	2.5	21×21	1.31
Identification (challenging)	80	5	64×64	4

of the building and camera's principal axis, which is given as

$$\rho_v = \rho_h \frac{H_I}{W_I} cos(20^\circ), \tag{2.4}$$

where ρ_h, ρ_v are pixel densities, and W_I, H_I are image dimensions in the horizontal and vertical directions, respectively. For example, for an HD video (1920x1080), required value of ρ_v to recognize a face is 1.3214 px/cm (good condition) or 2.6429 px/cm (challenging condition).

Regarding face recognition by machine algorithms under different poses and illumination conditions, T. Marciniak et al. [121] performed experiments to determine the required minimum resolution. They used MUCT data set and a PCA/LDA based face recognizer. It was found that under a pose variation of 21° and the down-sampling of the faces until 21×21 pixels, accuracy remains above 70%. This was just about 10% less than the accuracy of the faces having 256×256 pixels. For the illumination variations, they found that the required resolution is 64×64 pixels for the same accuracy, i.e. greater than 70%. These results and the corresponding pixel densities are also stated in Table 2.2.

Recently, C. Lu and X. Tang [122] showed that their face recognizer called GaussianFace achieved superior performance than humans on Labelled Faces in the Wild (LFW) data set which contains high-resolution faces collected under uncontrolled environment. However, the effect of low-resolution is not studied.

2.5 Limitations of the State-of-the-Art

This thesis is about protecting a bystander's privacy after capturing the images and videos; therefore, it leaves out pre-processing privacy filters from comparative analysis. Moreover,

the thesis only focusses on the main-identifiers especially faces; therefore, it does not consider works related to context protection. Thus, based on the three properties required for recreational photography stated in Chapter 1 (i.e. attacks robustness, minimal spatio-temporal distortion and computational complexity), the thesis presents a more detailed comparison of post-processing privacy filters. A summary of this comparison is given in Table 2.3, which includes a representative of reversible adaptive [20], reversible non-adaptive [19] and irreversible non-adaptive privacy filters [41]. The rest [4, 13, 21, 47] and proposed are irreversible & adaptive filters. Compared to Table 2.1, Table 2.3 only includes post-processing, sensitive-region oriented, autonomous privacy filters which can be adaptive, non-adaptive, reversible and irreversible.

2.5.1 Spatio-temporal Distortion

The distortion control mechanism in the state-of-the-art adaptive privacy filters is fully or partially manual [13, 21, 47], for example by manually selecting the kernel sizes for the given images [13] or videos [20, 21]. Similarly, when the filter intensity from the centre to the boundary of a detected face is automatically decreased, the kernel size at the centre is still manually selected for the given video [47]. In addition, the 2D isotropic kernels in state-of-the-art adaptive filters [13, 21, 47] can degrade a sensitive-region more severely when the sensitive-region resolution is different in the horizontal and vertical direction as may be typical in oblique images and videos. Thus, it is required to exploit the different horizontal and vertical resolutions, and use anisotropic kernels and a fully automatic adaptivity based on the resolution of the detected face. Moreover, state-of-the-art adaptive privacy filters based on pseudo-random numbers, e.g. scrambling or warping often change their parameters without considering associated temporal artefacts such as flicker generated by the abrupt changes in the intensity values of protected frames.

2.5.2 Robustness

We compare robustness of reversible adaptive filters [20], reversible non-adaptive filters [19], irreversible adaptive filters [41] and irreversible non-adaptive filters [4, 13, 21, 47] against brute force, naïve, parrot and reconstruction attacks.
Both reversible adaptive filters and reversible non-adaptive filters are robust against a parrot and a reconstruction attack, but they are prone to brute-force, spatial-domain [81, 81] and

Table 2.3: Post-processing privacy filters. KEY – DCT-S: Discrete Cosine Transform Scrambling; PICO: Privacy through Invertible Cryptographic Obscuration; GARP: Gender, Age and Race Preservation; SVGB: Space Variant Gaussian Blur; ODBVP: Optimal Distortion-Based Visual Privacy; AGB: Adaptive Gaussian Blur; AHGMM: Adaptive Hopping Gaussian Mixture Model. Adaptive control modulates the strength of a privacy filter.

			DCT-S [20]	PICO [19]	GARP [41]	UAS-VPG [4]	Cartooning [21]	SVGB [47]	ODBVP [13]	AGB (Chapter 3) [29]	AHGMM (Chapter 4) [123, 124]
Distortion	adaptive control	image based	✓				✓	✓	✓		
		navigation sensors								✓	✓
	spatial 2D kernel	isotropic					✓	✓	✓		
		anisotropic								✓	✓
	temporal correlation										✓
Robustness	to brute force attack				✓	✓	✓	✓	✓	✓	✓
	to naïve attack		✓	✓	✓	✓	✓	✓	✓	✓	✓
	to inverse filter attack				✓		✓				✓
	to super-resolution attack		✓	✓	✓						✓
	to parrot attack	with detectors			✓						
		without detectors	✓	✓							✓
Computational simplicity						✓		✓	✓	✓	

frequency-domain [101] attacks. In contrast, irreversible non-adaptive filters like GARP [41] are robust against the parrot and reconstruction attacks. However, their suitability highly depends upon sophisticated visual detectors such as pose, gender, race and age detectors. The efficiency of these detectors depends on face resolution, face pose and illumination conditions that are critical challenges in airborne photography. In contrast, irreversible adaptive filters [4, 13, 21, 47] do not depend on any sophisticated visual detectors, but they are prone to the parrot and reconstruction attacks.

2.6 Summary

This chapter stated the definition of privacy in visual data and formally defined a privacy filter. It reviewed and discussed the state-of-the-art privacy protection filters proposed for airborne as well as CCTV cameras by categorising them into pre-processing and post-processing privacy filters. It is found that pre-possessing are based either on access control which is insufficient in case of oblique optics or detecting every sensitive-region before capturing each frame which is a great challenge especially in an outdoor environment. Also, these filters are non-adaptive (i.e. mostly provide blank out) and therefore cannot control the distortion strength. In contrast, post-processing privacy filters are not based on access control and can provide high fidelity. However, adaptive post-processing privacy filters are prone to the parrot and reconstruction attacks, while non-adaptive post-processing privacy filters use sophisticated detectors and therefore depends upon their detection accuracy. Finally, the thesis highlights the research gap of developing robust privacy filters that do not require any sophisticated detector and also minimise spatio-temporal distortion.

Chapter 3

Privacy Design Space for Adaptive Privacy Filtering

This chapter defines a mechanism for *privacy design space exploration* that allows to automatically configure an adaptive privacy filter to protect a face. The mechanism uses the resolution of the detected face to determine when it is inherently protected. The mechanism exploits the auxiliary data from the on-board navigation sensors (GPS and IMU) to determine when a face is not inherently protected and then apply an adaptive privacy filter that distorts the face region depending upon its captured resolution. The block diagram of the proposed approach is shown in Figure 3.1.

The chapter is organised as follows: Section 3.1 discusses the concept of privacy design space and presents its analytical results. Section 3.2 describes the design of an adaptive privacy filter and shows example filtered faces. Section 3.3 presents experimental results. Finally, Section 3.4 summarizes the findings and limitations of the work presented in this chapter.

3.1 Privacy Design Space

Let an MAV fly at an altitude of h_1 meters. Let the principal axis $\boldsymbol{P}$ of its on-board camera be tilted by θ_P from the nadir direction $\boldsymbol{N}$ (see Figure 3.2). A value of $\theta_P \neq 0$ generates an oblique image. Each frame I_t at time t could contain U faces. However, for simplicity but without loss of generality the thesis only consider the case of $U = 1$. Let R_t represent the face region in the image, which is viewed at an angle θ_R. Finally, let h_2 be the height

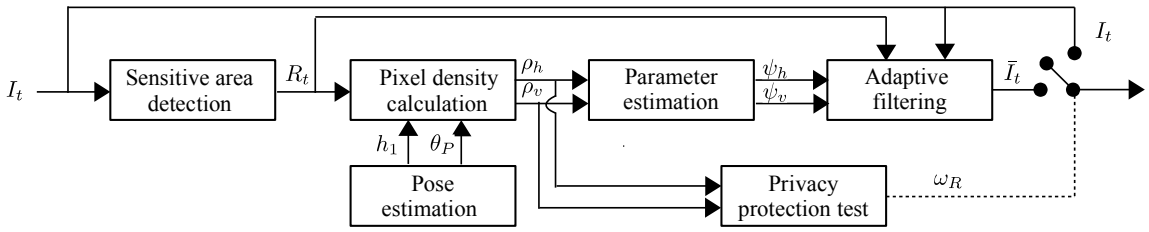


Figure 3.1: Block diagram of the proposed adaptive privacy filtering framework for airborne cameras.

of the face above ground.

Let the *pixel density* be defined as the number of pixels (px) per unit distance (cm) in a certain portion of the image I_t. Let ρ_h and ρ_v represent the pixel density (px/cm) around the centre C_R of R_t in the horizontal and vertical direction, respectively (see Figure 3.3). Our goal is to determine whether ρ_h and ρ_v are sufficiently large to facilitate the recognition of the identity of the person whose face is captured in R_t and then to apply an adaptive privacy filter in case recognition is considered possible.

Using reliable commercial-off-the-shelf differential GPS and IMU units, an MAV can estimate its position with an error that is smaller than $10-20$ cm [125]. The thesis assumes that on-board navigation sensor's data is available for each I_t with high accuracy and therefore use these auxiliary data to estimate h_1 and θ_P.

Given a face detection result R_t in I_t, θ_R is determined using C_R. For $\theta_P \neq 0$, a pixel of the image sensor covers an area of the 3D world of different size depending on its distance and orientation [126], [127]. Let p_h and p_v represent the physical dimensions of a pixel in the horizontal and vertical direction, respectively. If f is the focal length of the camera, the horizontal density ρ_h for a pixel around C_R can be determined as

$$\rho_h = \frac{f cos(\theta_R)}{p_h(h_1 - h_2)}, \tag{3.1}$$

and the vertical density ρ_v as

$$\rho_v \approx \frac{f cos(\theta_R) sin(\theta_R)}{p_v(h_1 - h_2)}. \tag{3.2}$$

Let the binary status variable $\omega_R \in \{0, 1\}$ define whether R_t is naturally protected ($\omega_R = 0$)

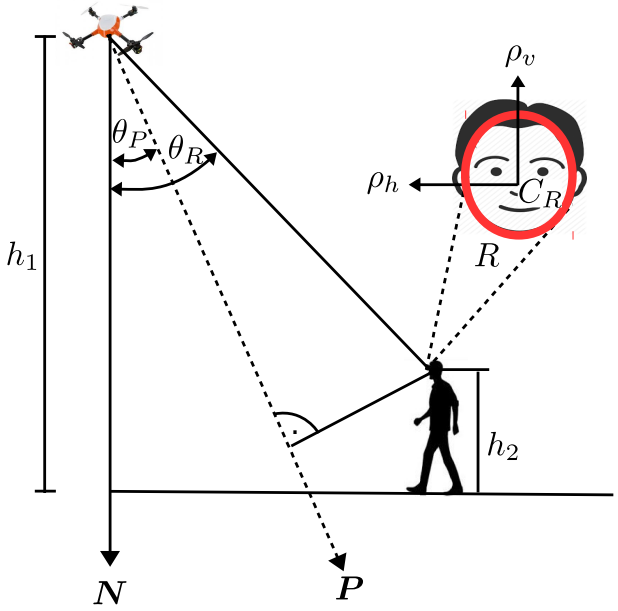


Figure 3.2: Capturing an image with an airborne camera at height h_1. The principal axis $\boldsymbol{P}$ of the camera is tilted by θ_P from the nadir direction $\boldsymbol{N}$. The sensitive-region R_t, a face at height h_2 above the ground, is viewed at an angle θ_R. The variables ρ_h and ρ_v represent the horizontal and vertical pixel density of R_t at its centre C_R in the captured image.

because of a low horizontal and vertical density, or not ($\omega_R = 1$):

$$\omega_R = \begin{cases} 1 & if \quad \rho_h > \rho_h^o \quad \& \quad \rho_v > \rho_v^o \\ 0 & otherwise \end{cases} \tag{3.3}$$

where ρ_h^o and ρ_v^o are experimentally defined thresholds. If $\omega_R = 0$, then the original frame I_t can be transmitted without any modifications.

Figures 4.3a and 4.3b show an example of ρ_h and ρ_v if we consider as camera a Canon EOS 5D Mark II in HD mode (1920×1080 pixels), whose sensor dimensions are 36 x 24 mm^2, thus $p_h = 18.75$ μm and $p_v = 22.22$ μm. We chose $h_2 = 1.7$ m and h_1 from 3 m to 150 m. Typical lenses have a focal length from 10 mm to 200 mm. Finally, θ_R is assumed to vary from 0° to 90°. If we accept to compromise on fidelity, we can globally filter I_t using ρ_h and ρ_v determined at $\theta_R = 45°$. This would lead in certain image-capturing conditions to an unnecessarily high level of blurring that reduces the fidelity of I_t.

Figure 4.3c depicts the boundary between the privacy sensitive space and the inherently

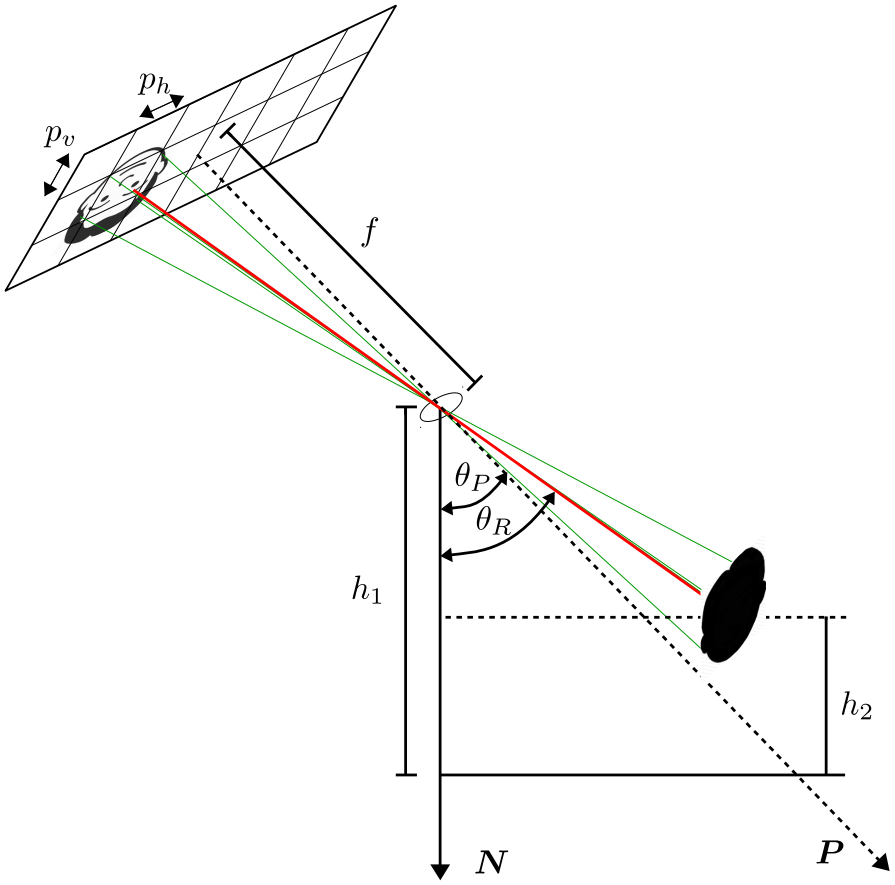


Figure 3.3: Mapping of a face at height h_2 to an image plane with oblique imagery. The height h_2 is considered between the ground and the centre of the face, C_R, which is seen from the back in the illustration.

protected space. The privacy sensitive space is determined by intersecting the individual segmentations based on ρ_h and ρ_v using the thresholds $\rho_h^o = 1$ px/cm and $\rho_v^o = 1$ px/cm.

3.2 Adaptive Privacy Filtering

When R_t is unprotected (i.e. the face is recognisable), we want to apply an adaptive privacy filter F_{Ω_j}, where Ω_j represents the parameter, so that the fidelity of the images can be increased with respect to the use of a fixed privacy filter. The value of Ω_j is such that the corresponding filtering with Ω_j turns ρ_h and ρ_v to be smaller than ρ_h^o and ρ_v^o, respectively.

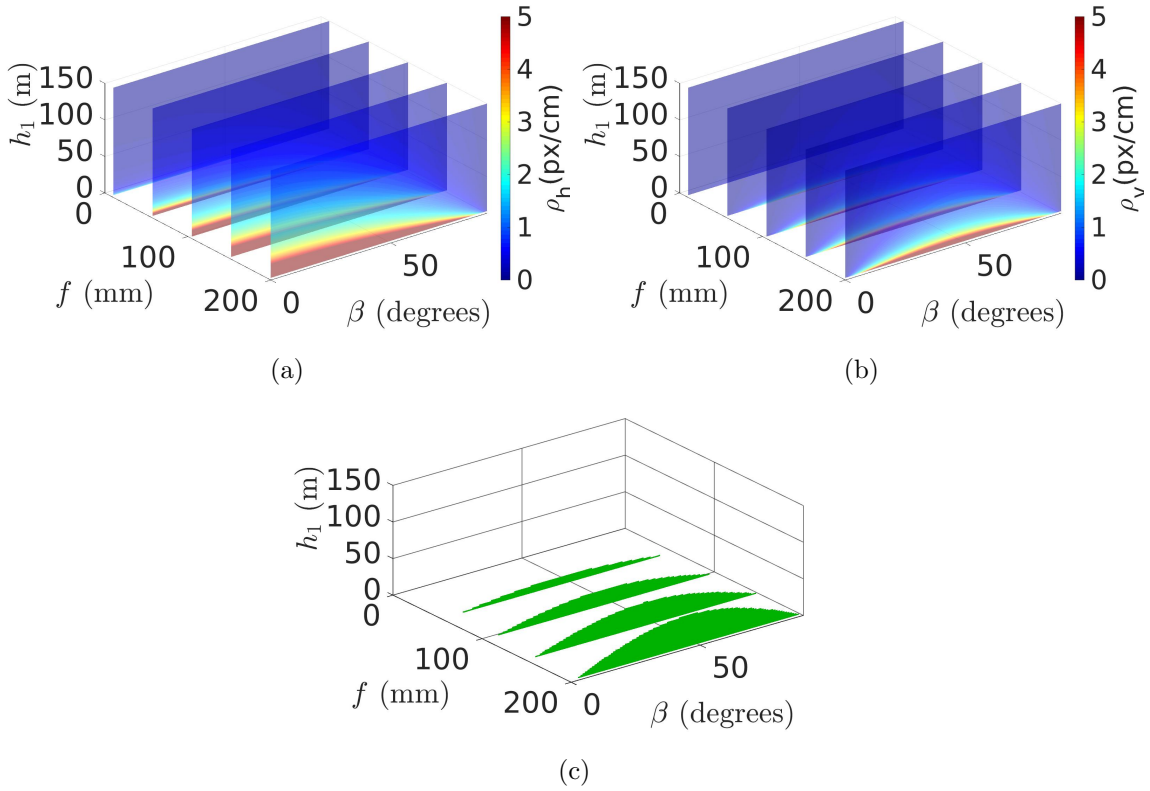


Figure 3.4: Horizontal and vertical pixel densities for EOS 5D Mark II in HD video mode at different heights, with different focal lengths and tilt angles. (a) Variation of ρ_h for different values of h_1, f and θ_R. The largest values of ρ_h occur at $\theta_R = 0°$. (b) Variation of ρ_v for different values of h_1, f and θ_R. The largest values of ρ_v occur at $\theta_R = 45°$. (c) Separation of the privacy sensitive space from the inherently protected space using $\rho_h^o = \rho_v^o = 1$ px/cm as segmentation threshold. The privacy sensitive space (indicated by the green slices) corresponds to the intersection of the segmentation performed in (a) and (b).

For example, in pixelation Ω_j is determined by the averaging kernel size [26], in blurring by the standard deviation of a Gaussian [28], in cartooning by the kernel size of a mean shift filter [21], in scrambling by the number of transform coefficients [31] and in warping by the relocation distance of pixels with respect to the calculated values of ρ_h and ρ_v [30].

In this work, we use the Gaussian blur that is widely used for outdoor imagery [87]. Therefore, F_{Ω_j} becomes a convolution operation between R_t and a Gaussian Point Spread Function (PSF). Basically, a 2D PSF $g(h, v)$, or impulse response, is the output of a filter when the input is a point source. In the discrete domain [128], it is given as $g(h, v) = \delta(h, v) * g(h, v)$, where $*$ is the convolution operation and

$$\delta(h, v) = \begin{cases} 1 & \text{if } h = v = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

In the case of Gaussian blur, $g(h, v)$ is an approximated Gaussian function of mean μ_j and standard deviation σ_j, and thus called a Gaussian PSF of parameter $\Omega_j = (\mu_j, \sigma_j)$. An approximated anisotropic Gaussian PSF is defined as

$$g(h, v) = \frac{1}{2\pi\sigma_h\sigma_v} e^{-\left(\frac{(h-\mu_h)^2}{2\sigma_h^2} + \frac{(v-\mu_v)^2}{2\sigma_v^2}\right)}, \tag{3.5}$$

where μ_h, μ_v, σ_h and σ_v are the mean and standard deviations of the Gaussian in the horizontal and vertical direction, respectively.

In a traditional Gaussian blur $\mu_j = 0$ [13, 47], and $\sigma_j \in \{\sigma_{jl} | l \in \mathbb{N}, \sigma_{jl+1} > \sigma_{jl}\}$ controls the distortion strength of F_{Ω_j} and provides pixel density $\rho_j \in \{\rho_{jl} | l \in \mathbb{N}, \rho_{jl+1} < \rho_{jl}\}$ in $\bar{R}_t$, respectively. As a higher σ_j results into lower ρ_j, we first find the minimum value called optimal parameter σ_j^o of σ_j that makes $\rho_j < \rho_j^o$. As a result, σ_j^o provides the minimum distortion in $\bar{R}_t$ while making it robust against the naïve-T attack (i.e. $P(\bar{R}_t | \mathcal{R}_\mathcal{G}) \rightarrow \epsilon$). Increasing σ_j beyond σ_j^o increases the distortion without improving the privacy level as the recogniser performance is already at the level of the random classifier. For a face captured from an MAV with pixel densities ρ_j, we calculate optimal parameters $\Omega_j^o = (\mu_j^o, \sigma_j^o)$ of a Gaussian PSF, where $\mu_j^o = 0$ and σ_j^o is estimated as follows:

A Gaussian PSF of standard deviation σ_j^o in the spatial domain is another Gaussian PSF of standard deviation $\acute{\sigma_j^o}$ in the frequency domain and both the Gaussian PSFs are related as

$$\acute{\sigma_j^o} = \frac{\rho_j}{2\pi\sigma_j^o}, \tag{3.6}$$

Figure 3.5: Comparison between various degrees of Gaussian blur to protect a facial image from an LDA face classifier. The numbers following $a*$ represent the value of ρ_h and ρ_v, respectively. The numbers following $b*$, $c*$, $d*$ represent the value of ψ_h and ψ_v, respectively. (a1-a7) Original images from [104] with decreasing pixel densities (from left to right). (b1-b7) Gaussian blur on (a1-a7) with $\rho_h^o = \rho_v^o = 0.5$ that results in slightly under-blurred images (i.e. the recognition rate is higher than that of a random classifier). (c1-c7) Adaptive Gaussian blur on (a1-a7) based on our proposed framework ($\rho_h^o = \rho_v^o = 0.4$) resulting in an adequate blurring of the faces for privacy preservation. With an adaptive Gaussian blur the kernel is selected depending on the pixel density for R_t. (d1-d7) Gaussian blur on (a1-a7) with $\rho_h^o = \rho_v^o = 0.3$ that results in slightly over-blurred images that, although they make the recognition rate equivalent to that of a random classifier, they unnecessarily decrease the fidelity of the facial images. (e1-e7) Fixed Gaussian blur of (a1-a7) using a *safe kernel* designed considering the highest possible pixel density in order to make the recognition accuracy of a classifier equivalent to that of a random classifier, irrespective of the pixel density for R_t, the face. This fixed Gaussian blur ($\psi_h = 121$, $\psi_v = 105$) significantly deteriorates the fidelity of lower resolution faces.

where $\acute{\sigma_j^o}$ is measured in cycles/cm, σ_j^o in px and ρ_j in px/cm. Let f_s represents the Nyquist frequency of ρ_j. Let $f_s^o < f_s$ is the highest spatial frequency component that we want to completely remove using a low pass filter, i.e. Gaussian blur. In other words, f_s^o is the Nyquist frequency of ρ_j^o, i.e. pixel density after filtering. Both ρ_j^o and f_s^o are related as

$$\rho_j^o = 2f_s^o. \tag{3.7}$$

As we are interested in removing frequency components beyond f_s^o, we can select $f_s^o = 3\acute{\sigma_j^o}$ because the amplitude response of a Gaussian PSF at three times of its standard deviation is very close to zero and multiplication (convolution in space domain) with such a Gaussian PSF will suppress frequencies larger than f_s^o. Substituting $f_s^o = 3\acute{\sigma_j^o}$ in Eq. 3.7, in the resulting relation Eq. 3.6 and finally rearranging gives the optimal standard deviation of Gaussian PSF as

$$\sigma_j^o = \frac{3\rho_j}{\pi\rho_j^o}. \tag{3.8}$$

Finally, the convolutional kernel ψ_j is determined by sampling the Gaussian function upto three times of σ_j^o as

$$\psi_j = 2\lceil 3\sigma_j^o \rceil + 1. \tag{3.9}$$

After convolving with a kernel of size ψ_j, the useful information is reduced to ρ_j^o (px/cm). As a result a frame with protected face region $\bar{I}_t$ is generated. Figure 3.5 shows sample results of adaptive Gaussian blurring, estimated standard deviations σ_h and σ_v, and kernel sizes ψ_h and ψ_v. From the figure, it is apparent how ψ_h and ψ_v decrease with decreasing ρ_h and ρ_v, respectively. This adaptive behaviour of ψ_h and ψ_v aims at maintaining the fidelity of the images. The required values of ρ_h^o and ρ_v^o depend on the expected ability of a face recogniser to discriminate identities.

In the next section we quantify the benefits of the proposed blurring approach.

3.3 Experimental Results

3.3.1 Setup

In this section we first study different target resolutions to determine the threshold values for separating inherently protected and unprotected spaces. We then analyse the trade-off between privacy and fidelity for adaptive and fixed privacy filters using standard face recognition algorithms by measuring their recognition performance with images from airborne

cameras. We use the LDA [117] and LBPH [116] algorithms for face recognition. The LDA face recogniser reduces the class dimension by maximising the inter-class to intra-class scatter ratio. In contrast, the LBPH face recogniser encodes a local structure instead of a full image to reduce the class dimension.

To measure fidelity, we apply the Structural Similarity Index (SSIM) and the PSNR. The SSIM measures the image quality by comparing the degradation in the structure of a protected image with the original [129], while the PSNR represents the power ratio of the original image with respect to the error introduced by protection.

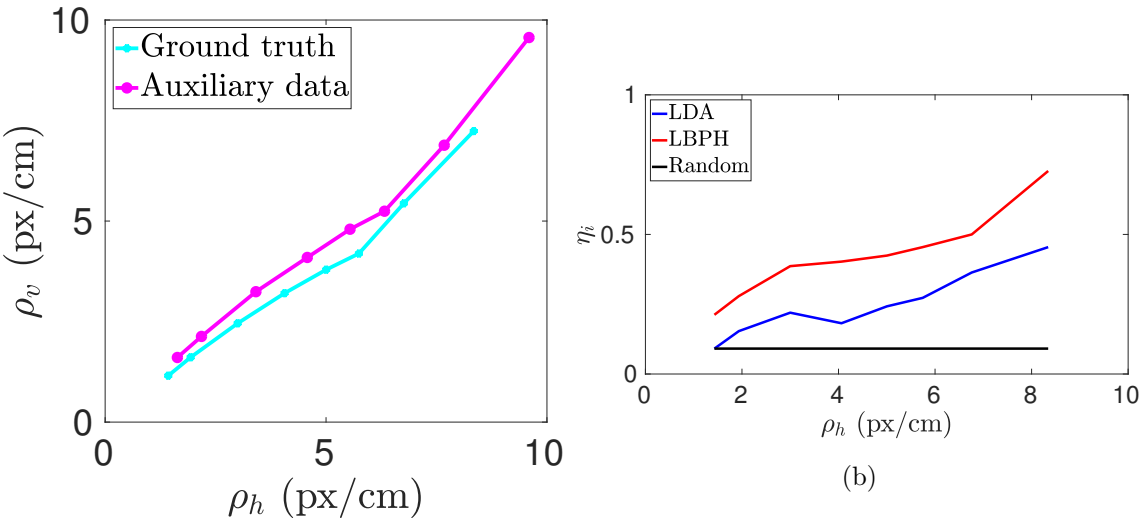
We use an outdoor data set emulating an MAV data set with the availability of auxiliary data, created with a camera placed at different heights and distances from faces [104]. In this data set, the principal axis of the camera is parallel to the ground, i.e. $\theta_P = 90°$. In order to compute the pixel densities for this particular setup, we modify Equations 3.1 and 3.2 as $\rho_h = \frac{f cos(90° - \theta_R)}{p_h d}$ and $\rho_v \approx \frac{f cos^2(90° - \theta_R)}{p_v d}$, where d represents the horizontal distance between the face and the camera.

We consider an image as privacy protected when the face recognition algorithms achieve an accuracy similar to that of a random classifier and therefore look for threshold values ρ_h^o and ρ_v^o resulting in a random classifier accuracy. The accuracy of a face recogniser corresponds to the rank-1 value of the cumulative match curve.

The data set contains 11 subjects and thus the accuracy of a random classifier for this data set is 0.091 (1/11). We chose data from 63 different positions for each individual resulting in a total of 693 test images (63×11).

To train the LDA and LBPH face recognisers, we used separate training images, which consisted of 11 images for each subject. For the detection of the face region, we annotated the images with the ground truth of the eye locations. As described in [130], we pre-processed all training and test images by (i) applying an affine transformation to compensate for scale and face rotation, (ii) using a bilateral filter to reduce noise and to compensate for light variations and finally (iii) masking to remove non-facial portions.

For the test images, we determined the values of ρ_h and ρ_v both by using auxiliary data (see Equations 3.1 and 3.2) and by manually counting the number of pixels on a face and normalising it by the standard face dimensions, i.e. the bitragion breadth of 15.9 cm and menton-crinion length of 21.9 cm [131]. As shown in Figure 3.6a, there is a small difference between the calculated pixel densities using these two methods with a Root Mean Square Error (RMSE) of 0.39 px/cm for ρ_h and 0.74 px/cm for ρ_v, respectively. The main reason for



(a) (b)

Figure 3.6: (a) The pixel density variation (ρ_h and ρ_v) of the 693 images in the data set based on auxiliary data and manually counting the pixels. (b) The achieved Rank-1 identification accuracy η_i of the LDA and LBPH face recognisers over the 693 raw images.

this error lies in our assumption of the identical height (h_2) and the identical face dimension for all 11 subjects.

3.3.2 Privacy Design Space

Figure 3.6b shows the recognition accuracy of LDA and LBPH over the original 693 test images as well as the response of a random classifier for reference. The LBPH face recogniser clearly outperforms the LDA face recogniser. However, the clear identification of the threshold values for the separation between the inherently protected and unprotected space is not possible with this data. For LBPH, the separating boundary lies *below* $\rho_h = 1.43$ px/cm (and corresponding $\rho_v = 1.15$ px/cm). Although LDA touches the random classifier level at $\rho_h = 1.43$ px/cm, this resolution cannot be considered as threshold with high confidence, because the accuracy of LDA also dropped at $\rho_h = 4.1$ px/cm and then again increased at $\rho_h = 2.99$ px/cm. Such behaviour may be due to the limited population size of the data set. Thus, the separating boundary lies *at or below* $\rho_h = 1.43$ px/cm (and corresponding $\rho_v = 1.15$ px/cm) for LDA.

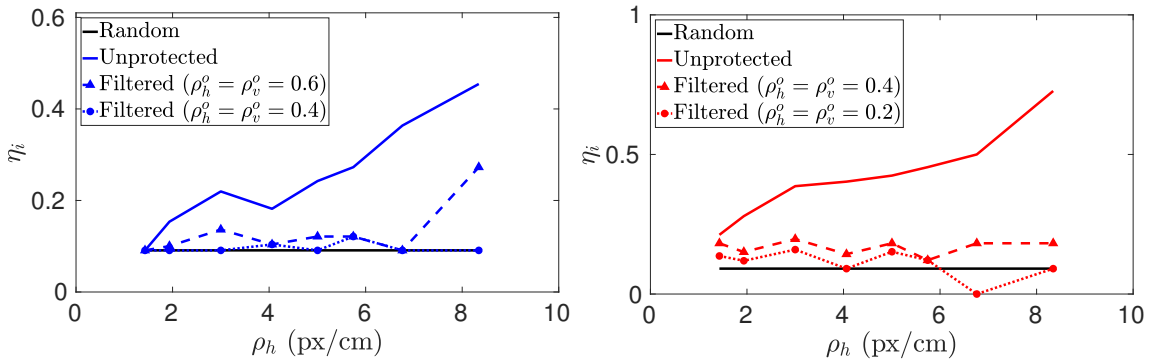


Figure 3.7: Rank-1 face identification accuracy η_i of adaptively filtered data with different threshold values ρ_h^o and ρ_v^o. (Left) LDA: The accuracy is similar to a random classifier at $\rho_h^o = 0.4$ px/cm, $\rho_v^o = 0.4$ px/cm. (Right) LBPH: The average response is close to a random classifier at $\rho_h^o = 0.2$ px/cm, $\rho_v^o = 0.2$ px/cm.

3.3.3 Adaptive Privacy Filtering

To explore the boundary further, we filter all test images with adaptive Gaussian blur to degrade the pixel resolution to the specified levels of $0.6, 0.4$ and 0.2 px/cm for both ρ_h^o and ρ_v^o. The kernel size of the Gaussian blur filter ($\psi_h \times \psi_v$) is computed as described in Section 3.2. We then determine the recognition accuracy over the degraded images. Figure 3.7 shows that the accuracy of LDA remains mostly around the random classifier accuracy of 0.091 at $\rho_h^o = 0.4$ px/cm and $\rho_v^o = 0.4$ px/cm. Thus, we define $\rho_h^o = 0.4$ px/cm and $\rho_v^o = 0.4$ as the boundary between the inherently protected and unprotected spaces for LDA. Although the accuracy of LBPH fluctuates, its average response is close to a random classifier at $\rho_h^o = 0.2$ px/cm and $\rho_v^o = 0.2$ px/cm. We therefore use $\rho_h^o = 0.2$ px/cm and $\rho_v^o = 0.2$ px/cm as the separating boundary for the LBPH.

Finally, we compare the fidelity of the adaptively filtered images with images filtered with a fixed safe kernel. Therefore, we apply adaptive Gaussian blurring with $\rho_h^o = 0.4$ px/cm and $\rho_v^o = 0.4$ px/cm and fixed Gaussian blurring with kernel size $\psi_h = 121 \times \psi_v = 105$ to all images and measure the SSIM and PSNR values. Figure 3.8 shows that adaptive filtering provides a higher fidelity in terms of SSIM and PSNR as compared to fixed safe filtering.

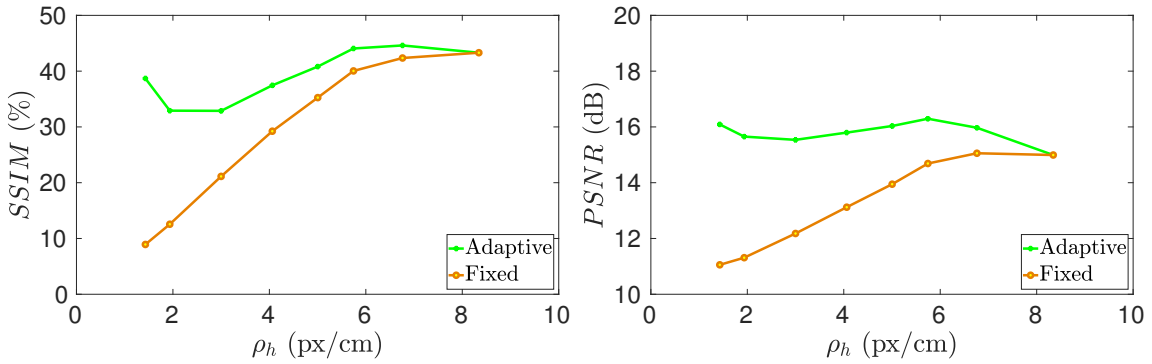


Figure 3.8: Fidelity achieved with adaptive and fixed privacy filtering over 693 images using the (top) Structural Similarity Index (SSIM) and the (bottom) Peak Signal to Noise Ratio (PSNR).

3.4 Summary

This chapter focused on face recognition by machine algorithms from airborne cameras and explored the design space to determine when a face in an airborne image is not recognisable. Moreover, when faces are recognisable, the chapter proposed an adaptive filtering that configured the strength of a privacy protection filter to improve the trade-off between privacy protection and usability of an aerial video.

There are few limitations of the presented work. First, the population size of the data set was small. Second, used face recognisers are not state-of-the-art. Third, adaptive Gaussian blur faces are prone to the parrot and reconstruction attacks. These limitations are addressed in Chapter 4.

Chapter 4

Robust Temporally-Smooth Seamless Privacy-Protection

This chapter presents a privacy filter AHGMM for drone photography/videography that improves the trade-off between privacy, fidelity and flicker. The proposed AHGMM distorts a face region with secret parameters to be robust to naïve, parrot and reconstruction attacks. Moreover, the filter minimises spatio-temporal distortion by adapting its parameters to the resolution of the captured face. In particular, an optimal Gaussian PSF that reduces the face resolution below a certain threshold is first selected. This protects the face against the naïve attack as well as maintains its resolution at a specified level. To prevent other attacks, supplementary Gaussian kernels in the selected Gaussian kernel are then inserted and their parameters are locally hopped using a PRNG so their estimation is difficult from the filtered face image. Finally, such protected faces are temporally averaged with decaying weights to minimise flicker.

The block diagram of the proposed AHGMM is shown in Figure 4.1. Face detection, pose estimation, pixel density estimation, privacy protection test and optimal Gaussian PSF blocks are explained in Section 3.1 and Section 3.2, while the rest of the blocks are explained in this chapter. For a single supplementary Gaussian PSF inside an optimal Gaussian PSF, the AHGMM is illustrated in Figure 4.2, while the pseudo-code is given in Algorithm 1.

The chapter is organised as follows: Section 4.1 states the objective of our proposed AHGMM filter. Section 4.2 describes the generation of the hopping kernels for privacy filtering. Section 4.3 gives the details of the local as well as global filtering and presents

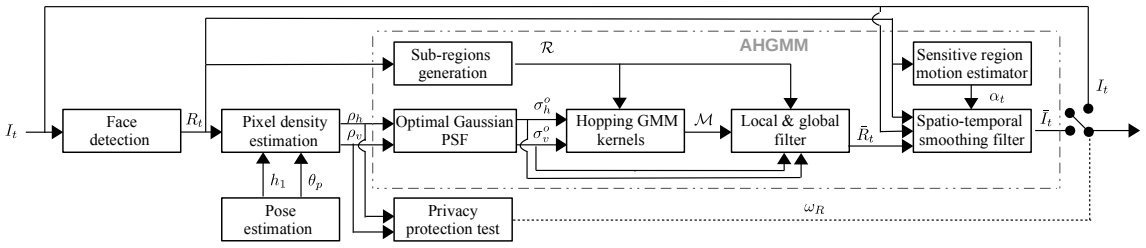


Figure 4.1: Block diagram of the proposed Adaptive Hopping Gaussian Mixture Model filter. KEY – ρ_h, ρ_v: number of pixels (px) per unit distance (cm) (pixel densities) of a sensitive-region R_t; h_1, θ_P: altitude and tilt angle of the camera used to calculate the pixel densities; ω_R: control signal generated from the pixel densities to decide when to protect R_t; $\mathcal{R}$: sub-regions of R_t; σ_h^o, σ_v^o: standard deviations for the hopping Gaussian mixture model $\mathcal{M}$ that filters $\mathcal{R}$ to generate a protected sensitive-region $\bar{R}_t$; $\bar{I}_t$: protected image created by spatio-temporal smoothing of $\bar{R}_t$.

sample filtered faces at different stages of the algorithm. Section 4.4 describes the details of a concatenated spatio-temporal smoothing filter that provides seamless protection and minimises flicker. Section 4.5 discusses the computational complexity of the algorithm. Finally, Section 4.6 presents the summary of the chapter.

4.1 Objectives

Our objective is to robustly protect a face against different attacks with a minimal spatio-temporal distortion, i.e. it aims at high fidelity and low flicker. Thus, our objective can be split into two competing targets: robust privacy protection and minimal spatio-temporal distortion.

4.1.1 Robust Privacy Protection

First of all, F_{Ω_j} should protect $\bar{R}_t$ such that it is not recognisable from humans and face recognition algorithms under the naïve-T attack (i.e. $P(\bar{R}_t|\mathcal{R}_\mathcal{G}) \rightarrow \epsilon$). In addition, the irreversible adaptive filter should ensure that the probability of correctly predicting the label of $\bar{R}_t$ is not increased in case of a parrot-T attack (i.e. $P(\bar{R}_t|\bar{\mathcal{R}}_\mathcal{G}) \rightarrow \epsilon$) as well as a reconstruction attack (i.e. $P(\hat{R}_t|\mathcal{R}_\mathcal{G}) \rightarrow \epsilon$ or $P(\hat{R}_t|\hat{\mathcal{R}}_\mathcal{G}) \rightarrow \epsilon$). Thus, the first objective of an

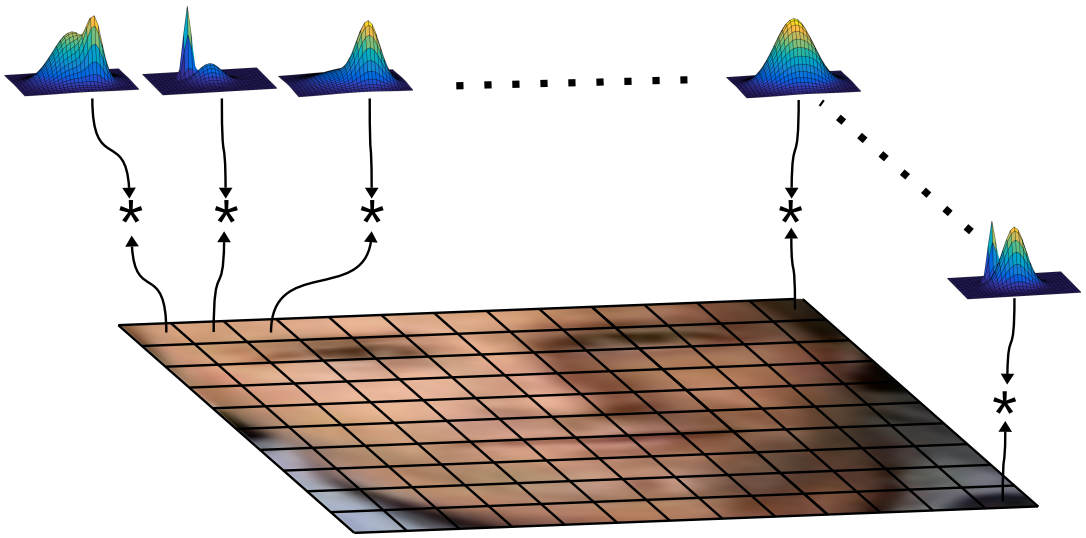


Figure 4.2: Visualisation of local filtering in AHGMM. The face region R_t is divided into Z sub-regions and each sub-region R_{tz} is convolved $(*)$ with a hopping Gaussian mixture model kernel M_z, which is made by an optimal Gaussian function and one (or more) supplementary Gaussian function added inside the optimal Gaussian function. While convolving with each sub-region of the face, the optimal and the supplementary Gaussian functions change their parameters, i.e. mean and standard deviation, which consequently changes the shape of the Gaussian mixture model based kernel.

irreversible privacy filter is as:

$$P(\bar{R}_t|\mathcal{B}) \rightarrow \epsilon. \tag{4.1}$$

In other words, the face image should be protected in such way that the probability $P(\bar{R}_t|\mathcal{B})$ of the protected face should not be better than a random classifier, irrespective whether the filtered or reconstructed face is compared against the unprotected, filtered or reconstructed gallery data sets. As the stated objective is dependent upon the recognition capability of a face recogniser, such objectives are heuristically solved for a given state-of-the-art face recogniser [42, 65, 96].

4.1.2 Minimal Spatio-Temporal Distortion

The second concurrent and competing target is that $\bar{R}_t$ should be protected with a minimal spatio-temporal distortion D. As a higher Ω_j results in lower ρ_j or higher D, an irreversible adaptive filter should optimally select parameter Ω_j to minimise spatial distortion. In

addition, if the resolution of R_t changes within a video, F_{Ω_j} should automatically adapt Ω_j to ensure a minimal spatial distortion. Finally, the selection of Ω_j in each frame should be in such a way that it also minimises any temporal distortion. Thus, the second objective of an irreversible privacy filter, without affecting the validity of Eq. 4.1, is as:

$$\bar{R}_t = F_{\Omega_j^*}(R_t), \tag{4.2}$$

where Ω_j^* is an ideal distortion parameter in each frame to minimise spatio-temporal distortion. If $E_t(w,h) = F_{\Omega_j}(R_t(w,h)) - R_t(w,h)$, the ideal distortion parameter, Ω_j^*, should be derived as:

$$\Omega_j^* = \arg\min_{\Omega_j} \left(\frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \big(E_t(w,h) + \xi(w,h)\big) + \big(P(\bar{R}_t|\mathcal{B}) - \epsilon\big) \right). \tag{4.3}$$

4.2 Hopping GMM Kernels

Filtering a detected face region R_t with the optimal Gaussian PSF (see Section 3.2) defined by Ω_j^o would only protect R_t from a naïve-T attack but not from a parrot-T attack and a recontrution attack. To ensure that the probability of correctly predicting the label of $\bar{R}_t$ is not increased in case of the parrot-T attack (i.e. $P(\bar{R}_t|\bar{\mathcal{R}}_\mathcal{G}) \rightarrow \epsilon$) as well as the reconstruction attack (i.e. $P(\hat{R}_t|\mathcal{R}_\mathcal{G}) \rightarrow \epsilon$ or $P(\hat{R}_t|\hat{\mathcal{R}}_\mathcal{G}) \rightarrow \epsilon$), Ω_j^o is secretly modified to Ω_j^* while generating $\bar{R}_t$ so that an adversary is unable to accurately reconstruct face region $\hat{R}_t$, or even generate $\hat{\mathcal{R}}_\mathcal{G}$ and $\bar{\mathcal{R}}_\mathcal{G}$. For this purpose, a set $\mathcal{R}$ is generated which consists of Z sub-regions in such a way that each sub-region covers a small area of R_t:

$$\mathcal{R} = \Big\{ R_{tz} | z \in [1, Z] \Big\}. \tag{4.4}$$

The size of R_{tz} (in pixels) affects the total number of sub-regions Z per face region R_t, which influences its privacy level. Smaller values of Z (bigger sub-regions) result in a reduced distortion.

After finding $\Omega_j^o = (\mu_j^o, \sigma_j^o)$ and generating $\mathcal{R}$, we make a hopping mixture of Gaussians for each sub-region, i.e. we pseudo-randomly change Ω_j^o to Ω_j^* for each R_{tz}. Moreover, we select supplementary Gaussian PSFs inside this optimal Gaussian PSF and vary their parameters, which are restricted by σ_j^o (lines 9-17 in Algorithm 1).

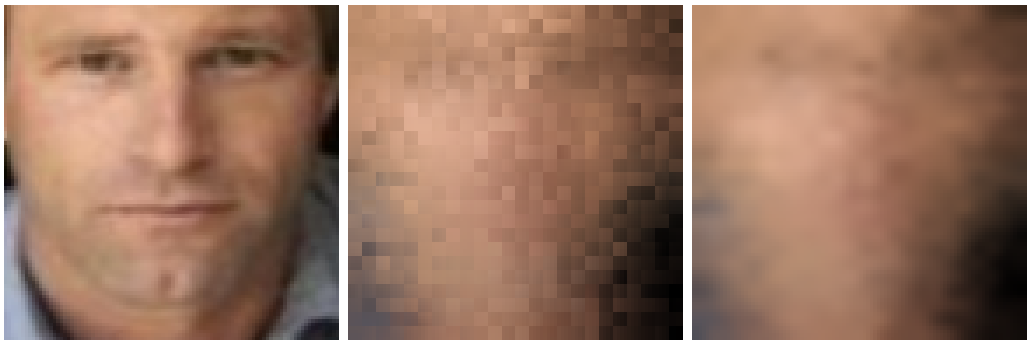
A set $\mathcal{X}$ containing the parameters of the Gaussian PSFs is built based on Z and the supplementary Gaussian PSFs for each sub-region, and is represented as

$$\mathcal{X} = \Big\{ (\mu_{jm}, \sigma_{jm})_z | z \in [1, Z], j \in \{h, v\}, m \in [0, M] \Big\}, \tag{4.5}$$

Algorithm 1 AHGMM

Input: I_t *unprotected image*
R_t *detected face region*
ρ_j *pixel density*, where $j \in \{h, v\}$
Output: $\bar{I}_t$ *protected image*

procedure FILTERAHGMM(I, R, ρ_h, ρ_v)
  for $j = h : v$ **do**
    $\mu_j^o \leftarrow 0$
    $\sigma_j^o \leftarrow \frac{3\rho_j}{\pi\rho_j^o}$
  end for
  $\mathcal{R} \leftarrow Z$ `sub-regions of` R_t
  for $z = 1 : Z$ **do**
    for $m = 0 : M$ **do**
      for $j = h : v$ **do**
        if $m = 0$ **then**
          $\mu_{jm} \leftarrow \pm\alpha_{jm}\sigma_j^o$
          $\sigma_{jm} \leftarrow (1 \pm \beta_{jm})\sigma_j^o$
        else
          $\mu_{jm} \leftarrow \pm\alpha_{jm}\sigma_j^o\gamma_{jm}$
          $\sigma_{jm} \leftarrow (1 \pm \beta_{jm})\sigma_j^o\gamma_{jm}$
        end if
      end for
      $X_z \leftarrow (\mu_{jm}, \sigma_{jm})_z$
      $G_{mz} \leftarrow$ `compute Gaussian functions`
      $\phi_{mz} \leftarrow$ `generate weights`
    end for
    $M_z \leftarrow$ `Gaussian mixture model`
    $\bar{R}_{tz} \leftarrow R_{tz} * M_z$
  end for
  $\bar{R}_t \leftarrow$ `apply global filter on` $\bar{\mathcal{R}}$
  $\bar{R}_t^s \leftarrow$ `apply spatio-temporal smoothing filter on` $\bar{R}_t$
  $\bar{I}_t \leftarrow$ `replace` R_t `with` $\bar{R}_t^s$ `in` I_t
  return $\bar{I}_t$
end procedure



(a) (b) (c)

Figure 4.3: Minimising blocking artefacts of spatially hopping Gaussian functions in AHGMM filter by a convolution with a global kernel. (a) Original image of 96×96 pixels from the LFW data set, (b) image after local filtering in AHGMM showing blocking artefacts and (c) image after the local filtering followed by the global filtering in AHGMM.

where M is the number of the supplementary Gaussian PSFs. The element $m = 0$ represents the modified optimal Gaussian PSF given by

$$\mu_{j0} = \pm\alpha_{j0}\sigma_j^o, \tag{4.6}$$

$$\sigma_{j0} = (1 \pm \beta_{j0})\sigma_j^o, \tag{4.7}$$

while the remaining elements (i.e. $m \in (0, M]$) belong to the supplementary Gaussian PSFs. These elements are calculated as

$$\mu_{jm} = \pm\alpha_{jm}\sigma_j^o\gamma_{jm}, \tag{4.8}$$

$$\sigma_{jm} = (1 \pm \beta_{jm})\sigma_j^o\gamma_{jm}, \tag{4.9}$$

where, $\alpha_{jm} \in [0, 1]$ and $\beta_{jm} \in [0, 1]$ are normalised pseudo-randomly generated numbers and control the local distortion in filtering. The variable $\gamma_{jm} \in (0, 1]$ controls the relative size of the supplementary Gaussian PSF w.r.t. the optimal Gaussian PSF.
After generating the parameters of the Gaussian PSFs, a set $\mathcal{G}$ representing 2D anisotropic-discretised Gaussian PSFs corresponding to $\mathcal{X}$ is created as

$$\mathcal{G} = \Big\{G_{mz} | z \in [1, Z], m \in [0, M]\Big\}, \tag{4.10}$$

where each G_{mz} is calculated (line 19 in Algorithm 1) as [132]

$$G_{mz} \approx A_{mz} e^{-\left(\frac{(h-\mu_{hmz})^2}{2\sigma^2_{hmz}} + \frac{(v-\mu_{vmz})^2}{2\sigma^2_{vmz}}\right)}, \tag{4.11}$$

where

$$A_{mz} = 1 \Big/ \sum_{(h,v)\in d} e^{-\left(\frac{(h-\mu_{hmz})^2}{2(\sigma_{hmz})^2} + \frac{(v-\mu_{vmz})^2}{2(\sigma_{vmz})^2}\right)}, \tag{4.12}$$

and

$$d = \left\{(h,v) \in \mathbb{Z}^2 : \left\lceil \frac{-\psi_h}{2} \right\rceil \leq h \leq \left\lceil \frac{\psi_h}{2} \right\rceil, \left\lceil \frac{-\psi_v}{2} \right\rceil \leq v \leq \left\lceil \frac{\psi_v}{2} \right\rceil \right\}, \tag{4.13}$$

with $\psi_j = 2\left\lceil 3\sigma_j \right\rceil + 1$. In order to develop a mixture model from the M discretised Gaussian PSFs of each sub-region, a set of weights ϕ is required. We again utilise PRNG to generate ϕ such that

$$\phi = \left\{\phi_{mz} | z \in [1, Z], m \in [0, M], \sum_{m=0}^{M} \phi_{mz} = 1\right\}. \tag{4.14}$$

Finally, a set of mixture models is generated for each sub-region (line 22 in Algorithm 1) as

$$\mathcal{M} = \left\{M_z | z \in [1, Z]\right\}, \tag{4.15}$$

where each element is calculated as

$$M_z = \sum_{m=0}^{M} \phi_{mz} G_{mz}. \tag{4.16}$$

4.3 Local and Global Filter

We have now Z discretised Gaussian mixture models in $\mathcal{M}$ for Z sub-regions of R_t. We locally convolve each sub-region R_{tz} (Eq. 4.4) with their respective M_z to make a protected sub-region $\bar{R}_{tz}$:

$$\bar{\mathcal{R}} = \left\{\bar{R}_{tz} | z \in [1, Z]\right\}, \tag{4.17}$$

where $\bar{R}_{tz} = R_{tz} * M_z$. Changing the convolutional kernel for each sub-region generates blocking artefacts (see Figure 4.3). To smooth these artefacts, we apply a global convolution filter (line 25 in Algorithm 1) with a Gaussian kernel of zero mean and standard deviation

$$\bar{\sigma}_j = \frac{\sigma^o_j}{Q_j}, \tag{4.18}$$

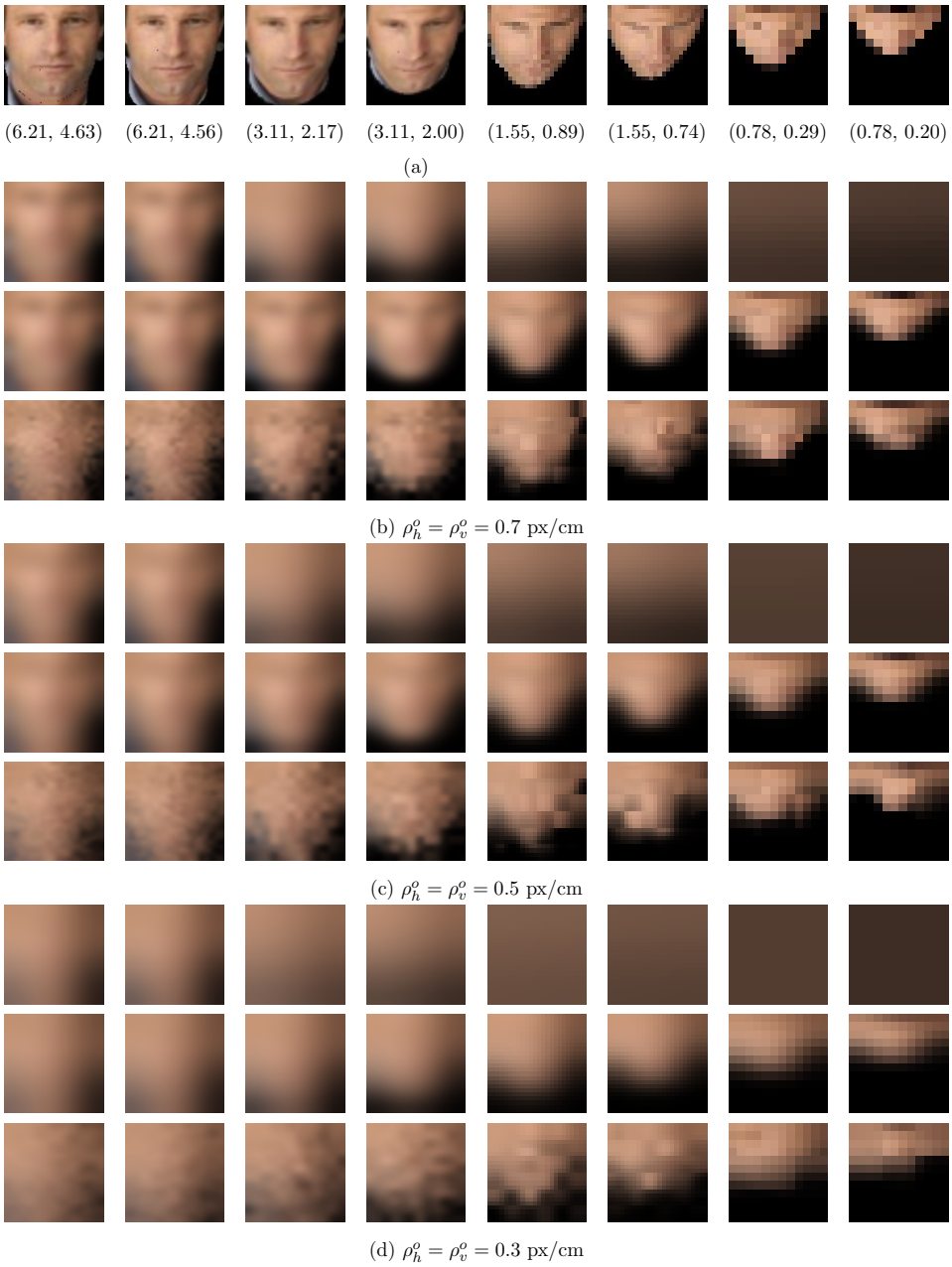


Figure 4.4: Visual comparison between fixed Gaussian blur (FGB), AGB [29] and AHGMM on the multi-resolution synthetically generated face data set. (a) Original images with pixel densities decreasing from left to right due different height and pitch angle. (.,.) indicates the horizontal and vertical pixel density in px/cm, respectively. (b-d) For various thresholds (ρ_h^o, ρ_v^o), results of FGB (first row), AGB (second row) and AHGMM filter (third row). For each threshold, FGB is selected w.r.t. the highest pixel density image in the data set. FGB does not adapt its parameters and therefore results into almost blanking out the image with smaller pixel density. In contrast, both AGB and AHGMM maintain high smoothness by varying their parameters depending upon the pixel densities of an image. Comparatively, AGB produces smoother images, while AHGMM filter creates blocking artefacts due to spatial switching of its parameters.

where Q_j represents the sub-region size in pixels. As a result, a spatially smoothed protected face $\bar{R}_t$ is developed. Figure 4.5 shows few sample images filtered by AHGMM at different thresholds.

4.4 Spatio-temporal Smoothing Filter

Directly replacing the original with the protected face region $\bar{R}_t$ may introduce strong boundary effects in a protectd frame $\bar{I}_t$. Moreover, in case of videography, the hopping Gaussian Mixture Model (GMM) for different sub-regions of a frame introduces flicker as the Gaussian mixture models are not temporally correlated and change from frame to frame independently.

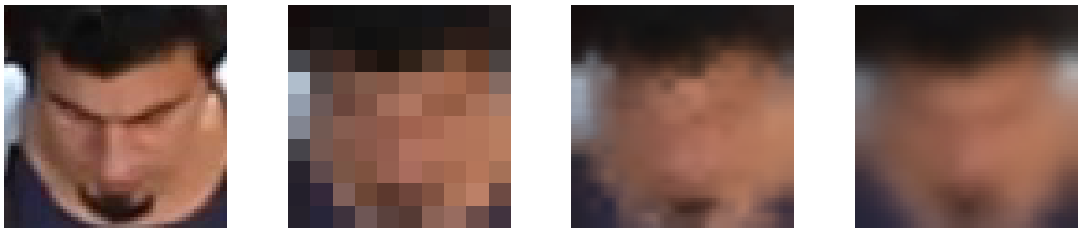
In order to mitigate these problems and generate a temporally smooth and seamlessly protected face $\bar{R}^s_t$, we blend the boundary of $\bar{R}_t$ and low-pass filter it as:

$$\bar{R}^s_t = \alpha_t[\alpha_s\bar{R}_t + (1-\alpha_s)R_t] + (1-\alpha_t)[\alpha_s\bar{R}_{t-1} + (1-\alpha_s)R_{t-1}], \quad (4.19)$$

where $\alpha_s \in [0,1]$ and $\alpha_t \in [0,1]$ are a spatial and a temporal weight, respectively. As a constant value of α_s blends $\bar{R}_t$ and R_t, but does not remove the sharp boundary between $\bar{R}_t$ and $\bar{I}_t$, we decrease α_s moving away from the boundary of the region. Moreover, a lower value of α_t increases smoothness but may introduce unpleasant delays in the video when the person moves. For this reason, to balance smoothness and delay, we adaptively select α_t depending on the motion of the face region, which is measured as displacement of the centres of $\bar{R}_t$ and $\bar{R}_{t-1}$.

4.5 Computational Complexity

The generation of a convolutional kernel is more complex in AHGMM than in the adaptive Gaussian blur filter (see Section 3.2). In fact, the latter only needs to compute a single Gaussian function, while AHGMM requires the computation of MZ Gaussian functions. Moreover, the adaptive Gaussian blur exploits the separability property of 2D convolutional kernels, i.e. $\psi = \psi_h * \psi_v$, to reduce the number of multiplications and additions from $W \cdot H \cdot |\psi_h| \cdot |\psi_v|$ to $W \cdot H \cdot (|\psi_h| + |\psi_v|)$ (W and H represent the width and height of R_t in pixels, respectively). Instead, AHGMM dynamically reconfigures the convolutional kernel after processing each sub-region and therefore requires exactly $W \cdot H \cdot |\psi_h| \cdot |\psi_v|$ multiplications and



(a) (b) (c) (d)

Figure 4.5: A face image protected with different privacy filters. (a) Original image (crop). Image protected with (b) pixelation, (c) AHGMM without temporal smoothing, and (d) AHGMM with temporal smoothing.

additions. Moreover, temporal smoothing filter requires additional $2 \cdot W \cdot H$ multiplications in case of videography.

4.6 Summary

This chapter presented an irreversible visual privacy protection filter that is based on spatio-temporal processing of faces. The proposed filter is based on an adaptive hopping Gaussian mixture model. Depending upon the captured resolution of a sensitive-region, the filter globally adapts the parameters of the Gaussian mixture model to minimise the spatial distortion, while locally hop them pseudo-randomly so that an attacker is unable to estimate these parameters. Finally, the boundaries of the protected sensitive-regions are spatially blended for a seamless insertion in a frame. Specifically for the videography, the filter concatenates spatially hopping GMM kernels with a temporally low pass filter to generate smoothly protected faces.

Unlike face-de-identification approaches ([12, 40, 41, 44, 91, 92, 95]), the proposed filter does not depend on a sophisticated detector (i.e. pose, facial expression, age, gender, race).

Chapter 5

Experimental Results

This chapter presents the experimental results to determine the validity of the proposed AHGMM algorithm to protect privacy as well as to minimise spatio-temporal distortion. The chapter separately discusses the experiments for photography and videography scenarios. For the photography experiments, the chapter uses a synthetically generated face image data set, while for the videography experiments, it exploits a small real airborne data set. For both photography and videography experiments, it assumes different levels of prior knowledge of an attacker and evaluates the privacy loss under a naïve attack, a parrot attack and a reconstruction attack. Moreover, it quantifies the corresponding fidelity degradation caused by the AHGMM and presents the trade-off analysis between privacy loss and fidelity. In the case of videography, the chapter also measures flicker subjectively and objectively, and presents the trade-off analysis between privacy loss, fidelity and flicker.
The chapter is organised as follows: Section 5.1 presents the experimental results for airborne photography. Section 5.2 discusses the experimental results for airborne videography. Finally, the chapter is concluded in Section 5.3.

5.1 Experimental Results for Photography

5.1.1 Setup

To the best of our knowledge, there is no large publicly available face data set collected from an MAV. We therefore generate face images as if they were captured from an MAV via geometric transformation and down-sampling on the LFW data set [133] (see Appendix A).

The LFW data set was collected in an unconstrained environment with extreme illumination conditions and extreme poses. We use the standard verification benchmark test of the LFW data set (12000 images of 4281 subjects), divided into 10-folds for cross-validation. Each fold contains 600 images of the same subject and 600 images of different subjects. We use the deep funnelled version of the LFW data set. Using these 12000 images, we synthetically generated 480,000 images emulating 40 different positions in space (i.e. 5 resolutions and 8 pitch angles).

We compare AHGMM against SVGB [47], Adaptive Gaussian Blur (AGB) (see Section 3.2) and Fixed Gaussian Blur (FGB), which uses a constant Gaussian kernel defined with respect to the highest resolution face. Thus, we estimate the kernel for FGB as in AGB for the face with 96×96 pixels at 0° pitch angle. For the SVGB filter, we divide the face into four concentric circles and reduce the kernel size by 5% while radially moving out between two consecutive regions as in [47]. Although the kernel for the innermost region was manually selected in the original work, we choose the anisotropic kernel as estimated by the AGB and convert it into an isotropic kernel for a fair comparison. We use a block size of 4×4 and $m = 1$ for the AHGMM.

To compare privacy filters, we measure the face verfication accuracy η_v (see Eq. 1.5) using OpenFace [134], an open source implementation of Google's face recognition algorithm FaceNet [135]. OpenFace uses a deep Convolutional Neural Network (CNN) as a feature extractor, which is trained by a large face data set (500k images). This feature extractor is applied on the training and test images for their representations (embeddings) which are used for verification/classification [135].

To measure distortion as in [21, 56], we apply the PSNR (see Eq. 1.1), the power ratio of the original image with respect to the filtered image.

We perform experiments with 480,000 images (consisting of 5 different resolutions and 8 different pitch angles) to determine the validity of the proposed AHGMM to protect the identity information of an individual. For this purpose, we analyse the effect of a naïve-T attack, a parrot-T attack, an inverse filter attack and a super-resolution attack. Moreover, we quantify the corresponding fidelity degradation caused by the AHGMM.

As AGB and SVGB do not use any secret key, we evaluate them only using their accurate parameters in the parrot-T, inverse filter and super-resolution attacks. In contrast, any of these attacks on AHGMM can be further divided into three sub-attacks: optimal kernel, pseudo AHGMM and accurate AHGMM. In the optimal kernel sub-attack, we assume that

Table 5.1: Attacks used to evaluate the privacy loss of the proposed AHGMM algorithm. Both the gallery faces and the probe faces can be protected or unprotected (naïve-BL). Moreover, the protected faces could be either unchanged or reconstructed (e.g. through an inverse-filter (IF) or super-resolution (SR)). Finally, any AHGMM attack could be further divided into three sub-attacks corresponding to the prior-knowledge of an attacker: optimal, pseudo and accurate.

				Gallery images			
				unprotected	protected		
					unchanged	reconstructed	
						IF	SR
Probe images	unprotected			naïve-BL	N/A	N/A	N/A
	protected	unchanged		naïve-T	parrot-T - optimal - pseudo - accurate	—	—
		reconstructed	IF	naïve-IF -optimal -pseudo -accurate	—	parrot-IF -accurate	—
			SR	naïve-SR -optimal -pseudo -accurate	—	—	parrot-SR -accurate

an attacker is able to estimate the parameters of the optimal kernel and applies the optimal kernel to the entire face. In the pseudo AHGMM sub-attack, we assume that the attacker knows the optimal kernel and randomly modifies the filter parameter for the N sub-regions. In the accurate AHGMM sub-attack, we assume that the attacker has access to the secret key and can decipher all filter parameters for the N sub-regions. As this prior-knowledge can be exploited for both probe and gallery images, we therefore evaluate AHGMM under 13 different scenarios stated in Table 5.1.

We assume that an attacker is able to determine the pitch angle of a protected face using the background information of an image captured from an MAV and can apply a geometric transformation to transform the gallery images at that pitch angle. Therefore, in all the following attacks, both the gallery and the probe images are at the same pitch angle which can be protected or unprotected depending upon the attack type. Moreover, we use the same resolution for both the gallery images and the probe images.

5.1.2 Naïve-T Attack

First of all, we perform a naïve-BL attack which shows the baseline face verfication accuracy when both the probe data set and the gallery data set are unprotected. The results of the naïve-BL attack are given in Figure 5.1. After that we perform a naïve-T attack in which the gallery images are unprotected, while the probe images are protected using FGB, SVGB, AGB and AHGMM. The results of this attack are given in Figure 5.2 at different thresholds ρ_j^o.

The naïve-BL attack shows that the accuracy η_v of our synthetically generated data set decreases with the decrease of the face resolution and with the increase in the face pitch angle. However, this trend vanishes at high pitch angles, i.e. 60° and 70°, where it shows slight randomness. Finally, for the low resolution faces (6×6 pixels), the accuracy does not show any effect of the pitch angle and slightly oscillates. Therefore, we consider 6×6 pixels inherently privacy protected and remove these images from the analysis of the privacy filters.

From the naïve-T attack, we are interested in finding the optimal threshold which defines the optimal kernel for AGB (see Eq. 3.8). It is clear from Figure 5.2 that the accuracy of the naïve-T attack decreases while decreasing the threshold. When the threshold reaches 0.5 px/cm, the difference between the accuracy achieved by AGB and a random classifier

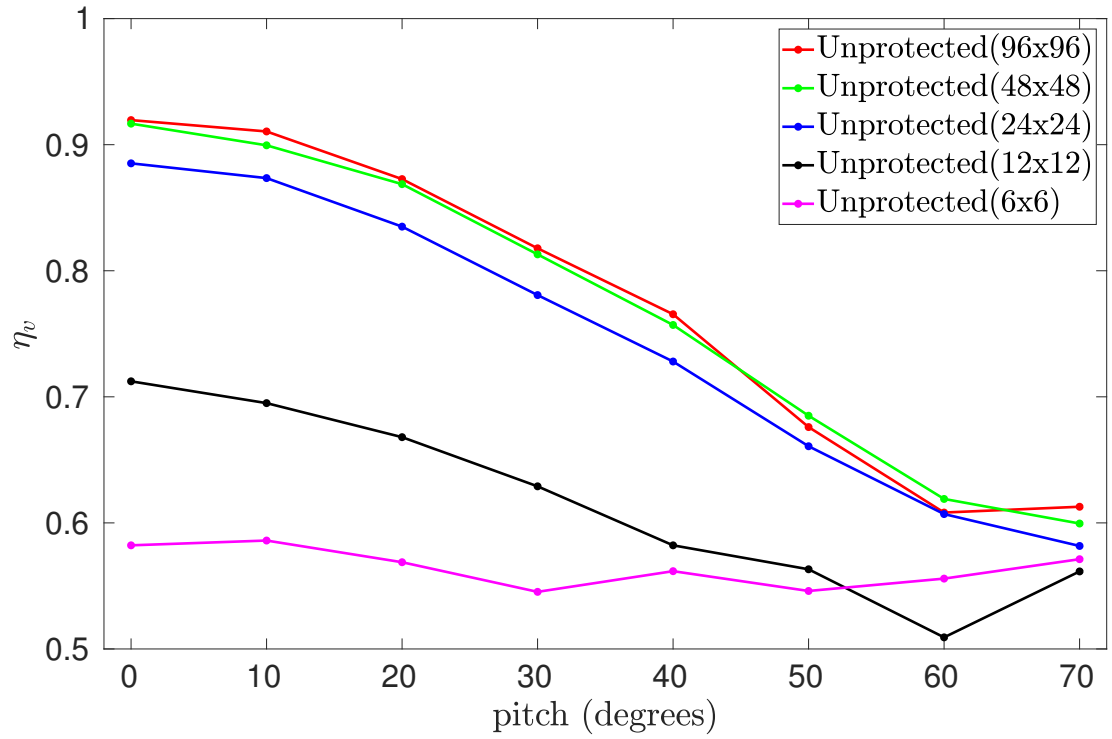


Figure 5.1: Face verification accuracy η_v of a naïve-BL attack on our synthetically generated face data set. In general, η_v increases with increasing the face size except at high pitch angles of 60 and 70 degrees where it slightly fluctuates randomly. For 6×6 pixels faces, η_v is the lowest and rather independent of the pitch angle.

($\eta_v = 0.5$) becomes very small except, unexpectedly, at high pitch angles. This difference further decreases at 0.4 px/cm and 0.3 px/cm. Thus, the optimal threshold defining the optimal kernel can be 0.5 px/cm, 0.4 px/cm and 0.3 px/cm. The later two thresholds decreases the accuracy negligibly but distort the images severely. Therefore, we decide to perform a trade-off analysis of the accuracy (under naïve, parrot attack and reconstruction attacks) and the distortion at these three thresholds.

At these three thresholds under the naïve-T attack, the accuracy of the AHGMM is higher as compared to the AGB. The main reason for this slightly higher accuracy is due to the under blurred sub-regions of the AHGMM filtered face as it hops its kernel below and above the optimal Gaussian kernel. In contrast, the accuracy of the SVGB is always lower than AGB and AHGMM. This is because SVGB uses an isotropic Gaussian kernel which deteriorates a face more severely as compared to the anisotropic kernel of the AGB and

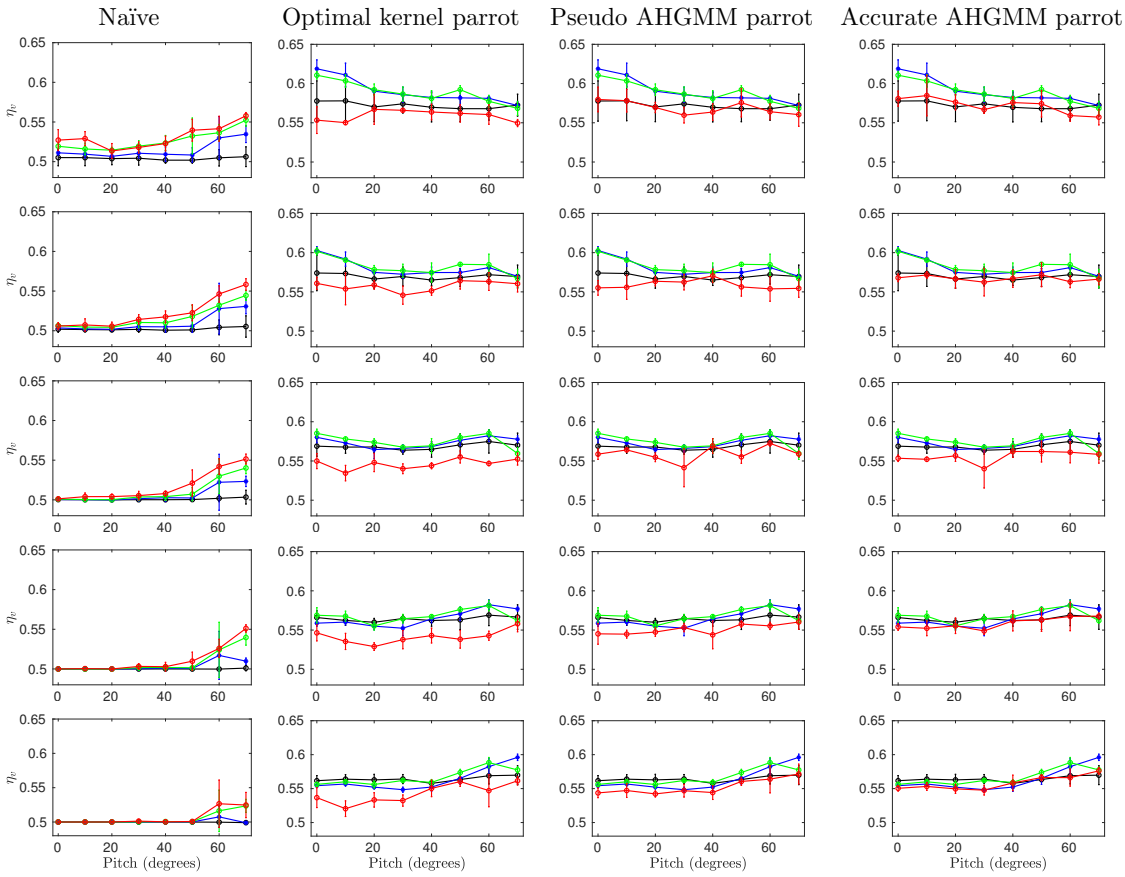


Figure 5.2: Face verification accuracy η_v achieved by naïve and parrot attacks on images protected by four different privacy protection filters at different thresholds ρ_j^o: first row: $\rho_j^o = 0.7$ px/cm, second row: $\rho_j^o = 0.6$ px/cm, third row: $\rho_j^o = 0.5$ px/cm, fourth row: $\rho_j^o = 0.4$ px/cm, fifth row: $\rho_j^o = 0.3$ px/cm. The filled marker shows the mean and the vertical bar indicates the standard deviation of η_v for the multi-resolution images (96×96, 48×48, 24×24, 12×12). Legend: AHGMM, AGB, SVGB, FGB. Under the naïve-T attack, AHGMM posses the highest η_v which converges towards $\eta_v = 0.5$ as the ρ_j^o is decreased and finally at $\rho_j^o \leq 0.5$ px/cm, the difference between η_v of AHGMM, AGB, SVGB and FGB becomes negligible, except unexpectedly at pitch angles 60° and 70° degrees. The parrot-T attack on AHGMM is divided into three sub-attacks: optimal kernel parrot-T attack, pseudo AHGMM parrot-T attack and accurate AHGMM parrot-T attack. In contrast to naïve-T attack, AHGMM provides the lowest η_v under any type of the three parrot-T attacks and this fact becomes negligible at $\rho_j^o = 0.3$ px/cm under accurate AHGMM parrot-T attack.

AHGMM filter. FGB possess the lowest accuracy at any threshold due to over blurring of all images except 96×96 pixels images at 0° pitch angle.

5.1.3 Parrot-T Attack

In the parrot-T attack, we filter both gallery and probe images and then evaluate the achieved accuracy. We study the parrot-T attack on AHGMM under three sub-attacks: optimal kernel parrot-T sub-attack, pseudo AHGMM parrot-T sub-attack and accurate AHGMM parrot-T sub-attack. The accuracy results of these sub-attacks are given in Figure 5.2 at different thresholds ρ_j^o, while Receiver Operating Curves (ROCs) for the accurate AHGMM parrot-T sub-attack at $\rho_j^o = 0.5$ px/cm are presented in Figure 5.3.

The parrot-T attack on state-of-the-art privacy filters increases the accuracy as compared to the naïve-T attack. Under the optimal kernel parrot sub-attack, our AHGMM shows the least accuracy improvement at any of the three thresholds. This is because the optimal kernel Gaussian blur is a spatially invariant blur that is not helpful in recognising spatially varying Gaussian blurred images, e.g. the AHGMM filtered images. Thus, our AHGMM provides the lowest accuracy against the parrot-T attack using the optimal kernel.

The pseudo AHGMM parrot-T sub-attack slightly improves the accuracy further as compared to the optimal kernel parrot-T sub-attack. The main reason is that both the gallery and the probe images are now filtered using spatially varying Gaussian blur. However, under the pseudo AHGMM sub-attack, the accuracy of AHGMM remains below the other three state-of-the-art privacy filters. Thus, our AHGMM provides the highest privacy protection even against the pseudo AHGMM parrot sub-attack.

Finally, the accurate AHGMM sub-attack improves the accuracy as compared to the optimal kernel and almost eqivalent to the pseudo AHGMM sub-attacks. Comparatively, even under the accurate AHGMM sub-attack, AHGMM performs better than FGB, AGB and SVGB at these three thresholds with the least improvement at $\rho_j^o = 0.3$ px/cm.

From the accurate AHGMM sub-attack, it is apparent that our AHGMM permanently removes the sensitive information from the face and an attacker can not recognise it with a high accuracy even when he/she has access to the secret key. This is in contrast to the reversible filters, e.g. encryption/scrambling based filters, which can reconstruct the original face after having the secret key. Thus, our AHGMM is robust against the brute-force attack.

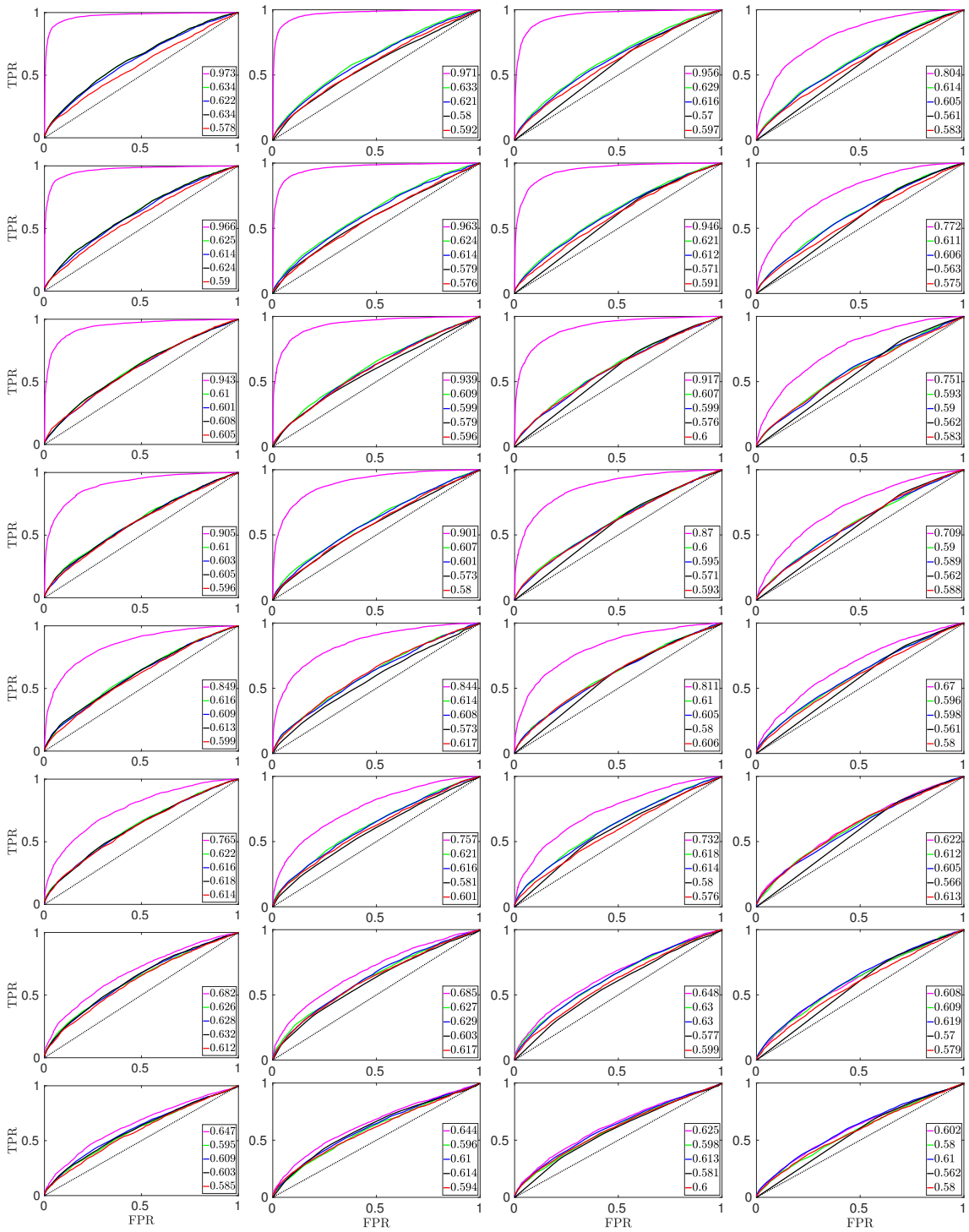


Figure 5.3: Receiver Operating Curves (ROCs) for the accurate parrot-T attack at threshold $\rho_j^o = 0.5$ px/cm. Each ROC is the mean of 10-curves generated by the 10-folds used for cross validation. Legend: Unprotected, AGB, SVGB, FGB, AHGMM. In each column, the pitch angle varies from 0° to 70° in 10° steps from top to bottom, while the image resolution remains same, i.e. first column (from left): 96×96 pixels, second column: 48×48 pixels, third column: 24×24 pixels and fourth column: 12×12 pixels. The legend values represent the Area Under Curve (AUC).

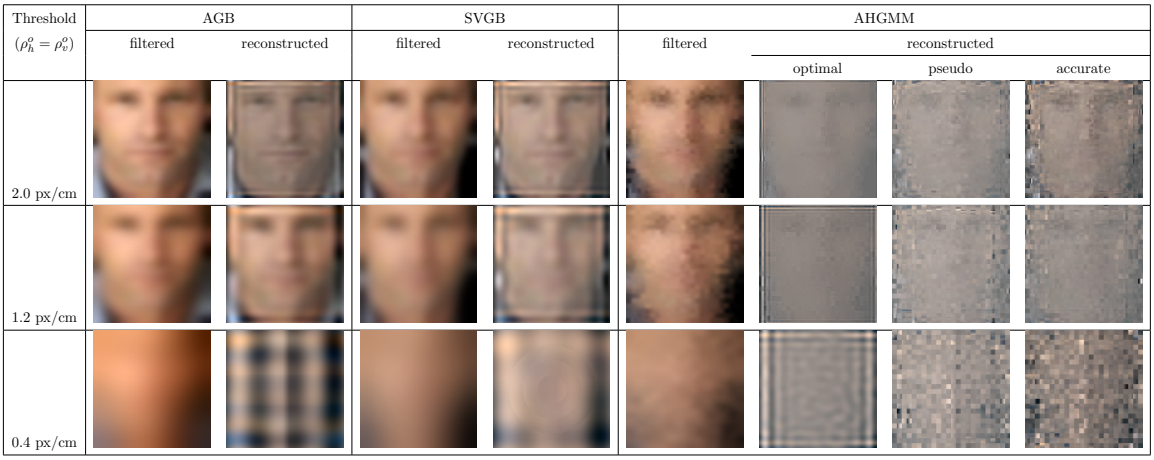


Figure 5.4: Inverse filtering of protected faces at different thresholds ρ_j^o. AGB and SVGB protected faces can be reconstructed by inverse filtering to some extent. Inverse filtering of AHGMM protected faces is hardly possible even if the hopping kernel parameters are known.

5.1.4 Inverse Filter Attack

In the inverse-filter (IF) attack, we reconstruct the probe images by deconvolving the protected face with an accurate or estimated kernel. We evaluate the IF attack under four sub-attacks: optimal kernel naïve-IF sub-attack, pseudo AHGMM naïve-IF sub-attack, accurate AHGMM naïve-IF sub-attack and accurate AHGMM parrot-IF sub-attack. Figure 5.4 depicts the effect of inverse filtering on selected sample images protected with AGB, SVGB and AHGMM. Figure 5.5 shows the achieved accuracies under the different sub-attacks at different values of ρ_j^o, while Figure 5.6 presents few ROCs for the accurate AHGMM parrot-IF sub-attack at $\rho_j^o = 0.5$ px/cm.

As can be seen in Figure 5.4, the face reconstruction quality decreases when the threshold increases (increasing the filter kernel) even if the filter parameters are known. This is true for both space invariant Gaussian blur (AGB) and linear space variant Gaussian blur (SVGB). The main reason is that the boundaries of the face start propagating towards the center of the face as the threshold is decreased. Thus, it becomes difficult to distinguish between reconstructed faces at the lower thresholds (see Figure 5.5).

In case of non-linear space variant blur (AHGMM), the reconstruction becomes more challenging even when the same hopping kernels are used as for the protection. The main

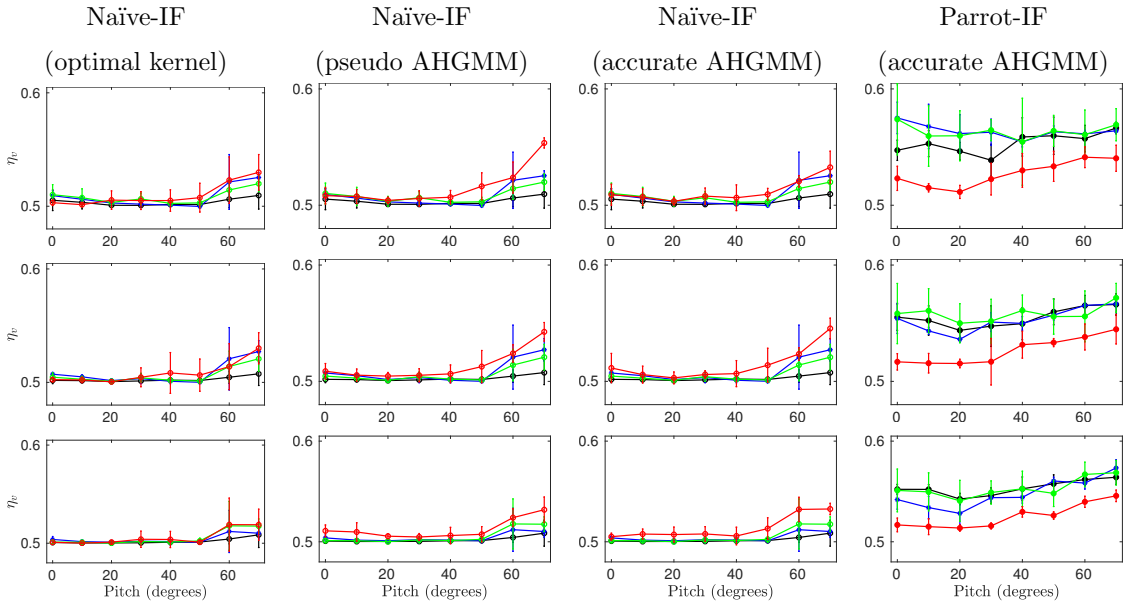


Figure 5.5: Face verification accuracy η_v achieved by an inverse filter (IF) attack on images protected by four different privacy protection filters at different thresholds ρ_j^o: first row: $\rho_j^o = 0.7$ px/cm, second row: $\rho_j^o = 0.6$ px/cm, third row: $\rho_j^o = 0.5$ px/cm. The filled marker shows the mean and the vertical bar indicates the standard deviation of η_v for the multi-resolution images (96×96, 48×48, 24×24, 12×12). Legend: AHGMM, AGB, SVGB, FGB. The IF attack is investigated under four sub-attacks: optimal kernel naïve-IF, pseudo AHGMM naïve-IF, accurate AHGMM naïve-IF and accurate AHGMM parrot-IF attack. The AHGMM achieves a slightly higher η_v under the naïve-IF attacks than the state-of-the-art filters, independently of the used threshold ρ_j^o. In contrast, AHGMM achieves the lowest η_v under the parrot-IF attack. As η_v is close to 0.5 under the naïve-IF attack for $0.5 \leq \rho_j^o \leq 0.7$ px/cm, we therefore do not perform experiments for $\rho_j^o < 0.5$ px/cm.

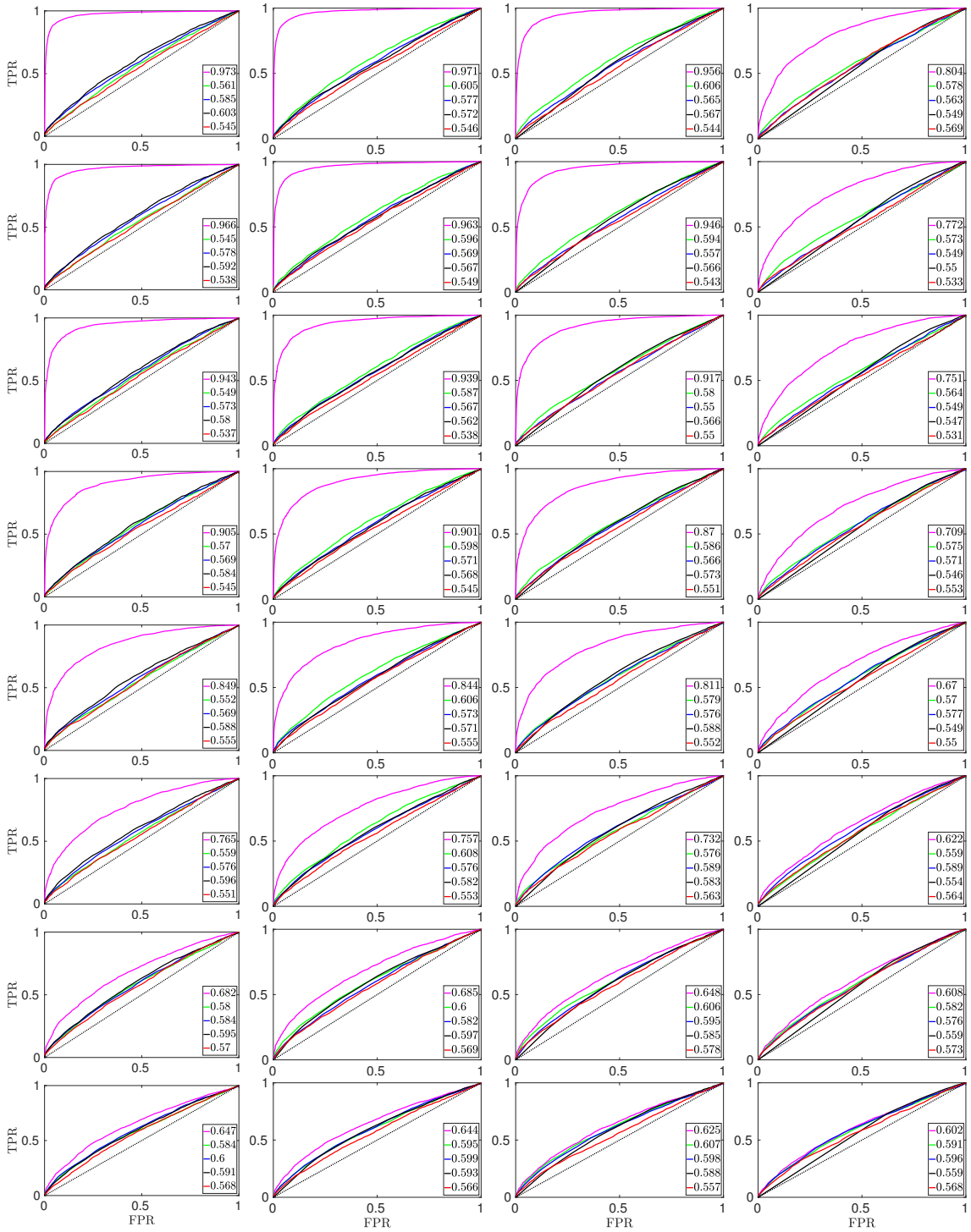


Figure 5.6: Receiver Operating Curves (ROC) for the accurate parrot-IF attack at threshold $\rho_j^o = 0.5$ px/cm. Each ROC is the mean of 10-curves generated by the 10-folds used for cross validation. Legend: Unprotected, AGB, SVGB, FGB, AHGMM. In each column, the pitch angle varies from 0° to 70° in 10° steps from top to bottom, while the image resolution remains same, i.e. first column (from left): 96×96 pixels, second column: 48×48 pixels, third column: 24×24 pixels and fourth column: 12×12 pixels. The legend values represent the Area Under Curve (AUC).

reason, in addition to the boundary propagation, is that while deconvolving a sub-region, the IF incorrectly treats the adjacent subregions as if they were filtered with the same kernel, thus not enabling it to reconstruct the original face (see Fig 5.4). Consequently, it becomes difficult to accurately predict the label of the reconstructed face.
In contrast to naïve-IF attacks, parrot-IF attack is more severe and increases significantly the accuracy, especially for AGB, FGB and SVGB. AHGMM also shows the accuracy improvement but less than AGB, FGB and SVGB, and is more robust to an inverse filter attack even when using an accurate secret key.

5.1.5 Super-resolution Attack

In this attack, we reconstruct the filtered probe images with SRCNN [22]. SRCNN first learns a mapping between the high-resolution images and their corresponding low-resolution version, and then applies this mapping to enhance the details of a low-resolution image. We learn the SRCNN mapping for $1,000,000$ iterations between the protected images (i.e. the low resolution) and their corresponding unprotected images (i.e. the high resolution) using the same data sets (91-images and Set5) as used in [22]. As learning of the mapping is a time consuming process, we investigate the super-resolution attack for a single point of our synthetic data set: 12000 images each with 96×96 pixels and $0°$ pitch angle.
We evaluate the super-resolution (SR) attack under four sub-attacks: optimal kernel naïve-SR sub-attack, pseudo AHGMM naïve-SR sub-attack, accurate AHGMM naïve-SR sub-attack and accurate AHGMM parrot-SR sub-attack. Tab. 5.2 summarises the achieved accuracies under the different sub-attacks, while Figure 5.7 presents the ROC for the accurate AHGMM parrot-SR sub-attack. Figure 5.8 depicts a visual comparison of the super-resolution reconstruction for three sample faces protected by AGB, SVGB and AHGMM filters.
For the space invariant Gaussian blur (AGB), it is apparent from Figure 5.8 that the SR attack can reconstruct the faces more effectively, even when the kernel size is quite high (i.e. $\rho_j^o = 0.5$ px/cm). Therefore, the faces protected by AGB achieves a higher accuracy (see Tab. 5.2). In contrast, faces protected by linear space variant Gaussian blur (SVGB) are difficult to reconstruct. The main reason is that the SR mapping becomes erroneous especially for patches which contain parts processed by different kernels. However, SR can effectively reconstruct patches where the Gaussian blur is locally invariant (e.g. compare

Table 5.2: Face verification accuracy η_v after a super-resolution attack on faces protected by adaptive Gaussian blur (AGB), space variant Gaussian blur (SVGB) and AHGMM at threshold $\rho_j^o = 0.5$ px/cm. The values of η_v are given as $\tilde{\mu}(\tilde{\sigma})$, where $\tilde{\mu}$ indicates the mean and $\tilde{\sigma}$ the standard deviation for the 10-fold cross validations. In the naïve-SR attack, the reconstructed probe faces are compared against the unprotected gallery images, while both the probe and the gallery images are super-resolved in the parrot-SR attack.

Attack type	AGB	SVGB	AHGMM
optimal naïve-SR	0.592 (0.012)	0.566 (0.016)	**0.515 (0.014)**
pseudo AHGMM naïve-SR	–	–	**0.520 (0.006)**
accurate AHGMM naïve-SR	–	–	**0.532 (0.018)**
accurate AHGMM parrot-SR	0.634 (0.015)	0.583(0.034)	**0.546 (0.018)**

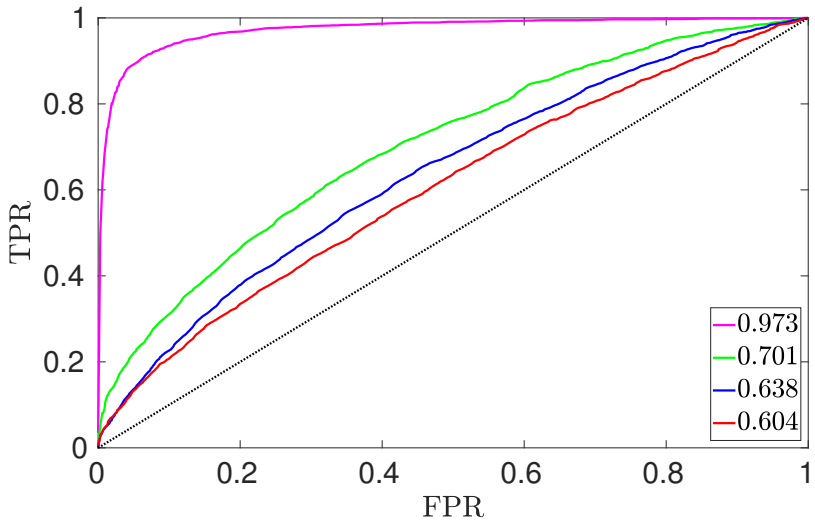


Figure 5.7: Receiver Operating Curve (ROC) for the accurate AHGMM parrot-SR attack at threshold $\rho_j^o = 0.5$ px/cm. Each ROC is the mean of 10-curves generated by the 10-folds used for cross validation. Legend: Unprotected, AGB, SVGB, AHGMM. This test is performed only for a single resolution (96×96 pixels) and pitch anfle ($0°$). The legend values represent the Area Under Curve (AUC).

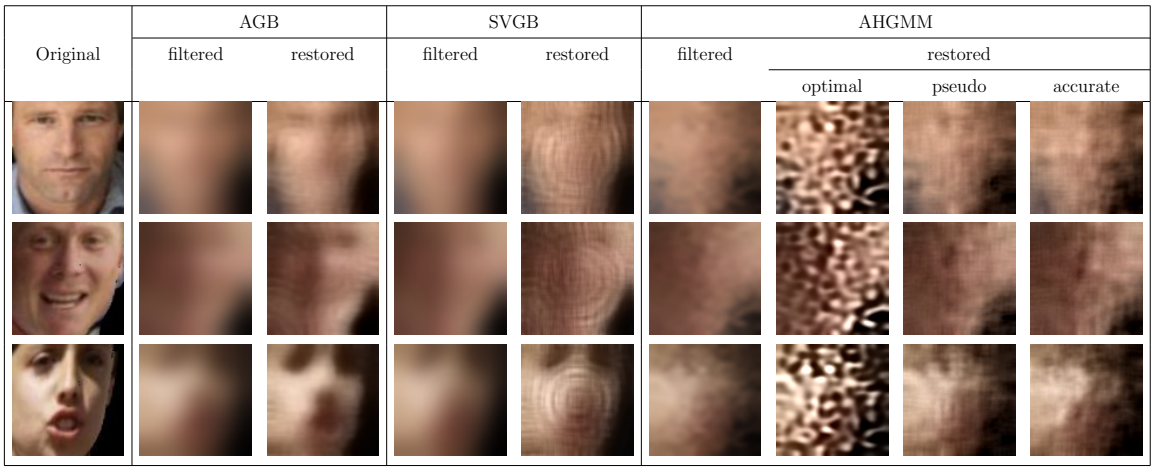


Figure 5.8: Visual comparison of reconstructed faces with super-resolution algorithm SR-CNN [22] for threshold $\rho_j^o = 0.5$ px/cm. Reconstruction performance deteriorates from AGB over SVGB to AHGMM protected faces.

the areas around eyes of the SVGB restored faces in Figure 5.8). The overall reconstruction is worse than for AGB and thus the achieved accuracy is lower.

Reconstruction by super-resolution is even more challenging for AHGMM protected faces. The main reason is that a single patch for learning the mapping contains several sub-regions each filtered with pseudo-randomly correlated Gaussian mixture models. Thus, the error in the learned SR mapping increases resulting in the lowest accuracy as compared to AGB and SVGB.

Similarly to parrot-IF attack, the accuracy improves for the parrot-SR attack where SR-reconstruction is also performed for the gallery images. Especially for AGB and SVGB, the similarity between (protected and reconstructed) gallery images and the (reconstructed) probe images increases. Thus, the accuracy increases. As for the other attacks, AHGMM is more robust to parrot attacks than AGB and SVGB, and achieves the lowest accuracy.

5.1.6 Distortion Analysis

We measure the distortion of the FGB, SVGB, AGB and AHGMM using PSNR. For a trade-off analysis between distortion and privacy, we plot the face verification accuracy against PSNR. The results of this trade-off analysis are presented in Figure 5.9.

AGB has the highest average PSNR values followed by SVGB, AHGMM and FGB. The

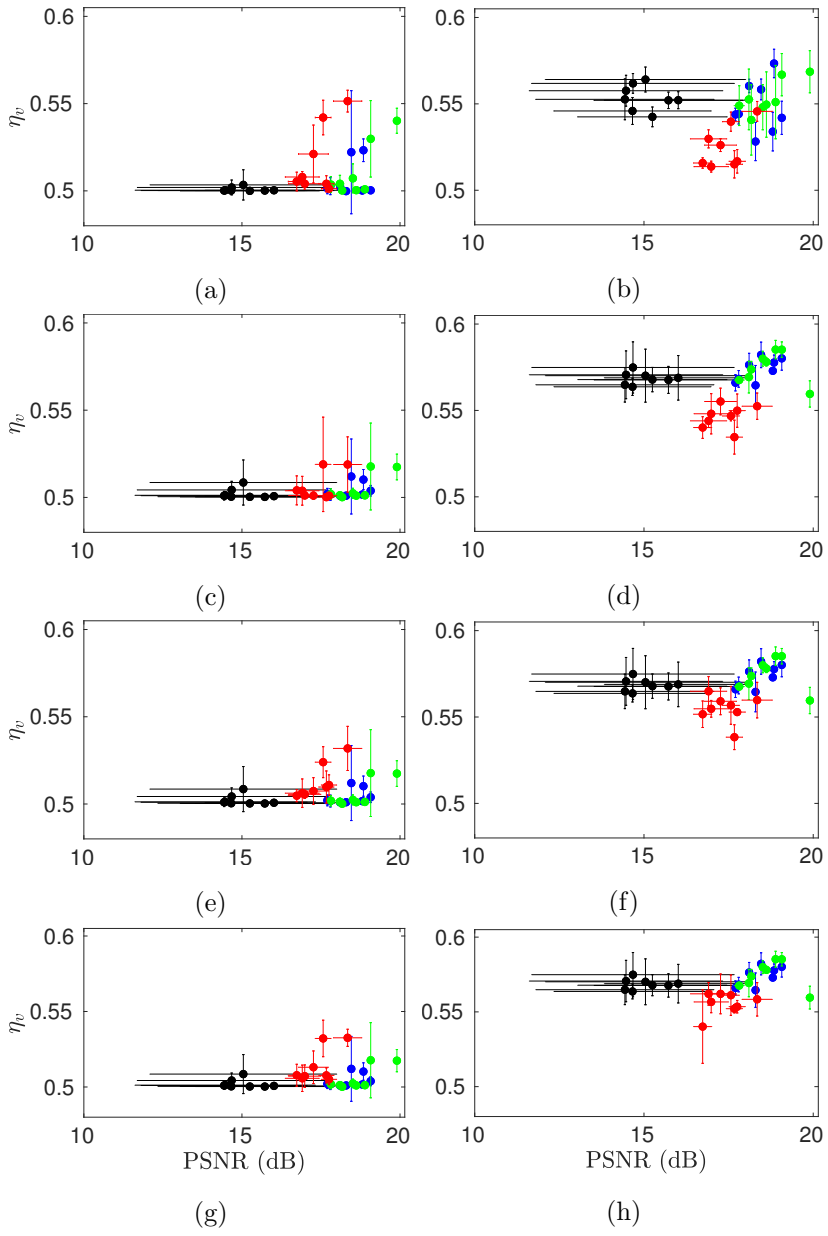


Figure 5.9: Trade-off analysis between the face verification accuracy η_v and the distortion provided by the different privacy filters under the naïve-T, parrot-T and inverse filter (IF) attacks at threshold $\rho_j^o = 0.5$ px/cm. The distortion is measured by the Peak Signal to Noise Ratio (PSNR). Legend: AHGMM, AGB, SVGB, FGB. Under the naïve-T attack, our proposed AHGMM possesses η_v almost equivalent to the state-of-the-art filter, but lowest under the parrot-T attacks. However, AHGMM has slightly lower PSNR as compared to AGB and SVGB, but much higher than FGB. (a) naïve-T attack, (b) accurate AHGMM parrot-IF attack, (c) optimal kernel naïve-IF attack, (d) optimal kernel parrot-T attack, (e) pseudo AHGMM naïve-IF attack, (f) pseudo AHGMM parrot-T attack, (g) accurate AHGMM naïve-IF attack and (h) accurate AHGMM parrot-T attack. For the last three naïve-IF and parrot-T attacks, the results of AGB, SVGB and FGB are the same and have been superimposed for the comparison. Please see Section 5.1.2, Section 5.1.3 and Section 5.1.4 for the details of the attacks.

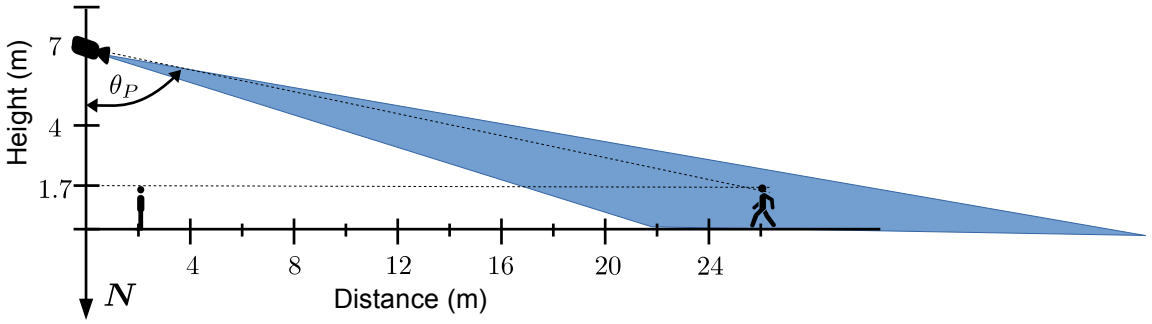


Figure 5.10: Setup for the collection of a probe video data set. The subject moves from a distance of 26 m to 2 m towards the camera, which is positioned at 7 m and at 4 m height. The variation of the pitch angle, θ_P, is about 20°-78° from the Nadir direction N.

main reason is that AGB uses a single anisotropic kernel instead of spatially linearly varying kernel used by SVGB. Although AHGMM also uses an anisotropic kernel like AGB, the spatial hopping phenomena of the Gaussian mixture model of the AHGMM results in high distortion (PSNR values) as compared to AGB and SVGB (see Figure 4.5). FGB has the highest distortion as it does not change its parameters depending upon the resolution of the face.

5.2 Experimental Results for Videography

In this section, to highlight the effect of temporal smooting (Section 4.4), the proposed AHGMM filter is split into two filters: AHGMM^p and AHGMM^v. The only difference between AHGMM^p and AHGMM^v is that the former does not use a temporal smoothing filter (same as used for the photography experiments presented in Section 5.1), while the later exploits a temporal smoothing filter.

5.2.1 Setup

We captured an Ultra-HD video probe data set with a GoPro5 camera mounted with a custom lens (25 mm) using the set up shown in Figure 5.10. For training, we captured an HD video gallery indoor data set with the built-in camera of a Lenovo K5 smart phone ensuring pitch angle variation of $10° - 90°$ degrees. There were 11 subjects in both data sets with only frontal faces. We extracted 7944 and 399 key-frames from the probe and gallery video

data sets, respectively, using the algorithm in [136], followed by manual post-processing to remove frames affected by motion blur. We preprocessed all key-frames by equalizing illumination, smoothing noise with a bilateral filter, aligning by an affine transformation using eye centres and finally applying elliptical masking to remove non-facial parts.

We compare AHGMM^v filter with AHGMM^p, AGB(see Section 3.2) and SVGB [47]. Both AHGMM^p and AHGMM^v are based on non-linear space variant Gaussian blur and use hopping Gaussian kernels, SVGB is a linear space-variant Gaussian blur that linearly reduces the kernel size while filtering a face, AGB is a space invariant Gaussian blur and uses a single Gaussian kernel. AGB is regarded as a flicker-free filter.

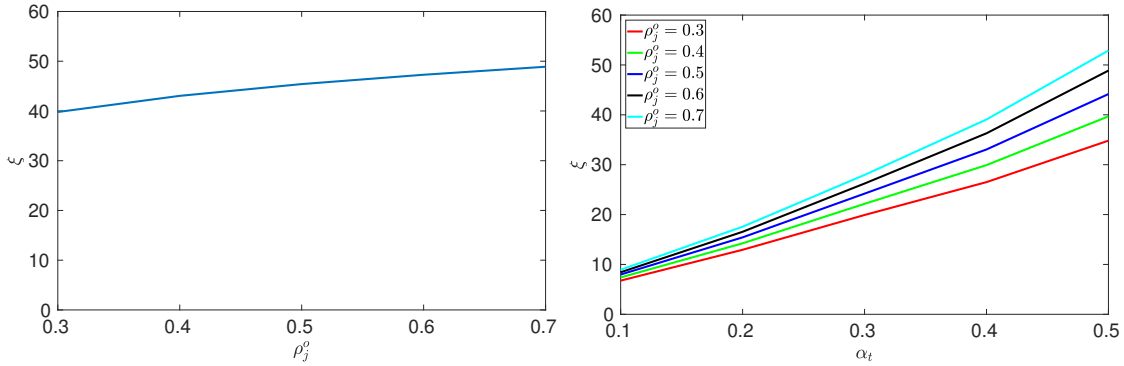
We evaluate all the filters under a naïve-T, parrot-T, pseudo naïve-SR and pseudo parrot-SR attacks. Training the super-resolution embeddings for each frame is a very time consuming process and also there is a very small difference between the pseudo and accurate super-resolution attacks (See Table 5.2), we therefore only perform pseudo naïve-SR and pseudo parrot-SR attacks for videography experiments. We use the SRCNN [22] super-resolution algorithm for naïve-SR and parrot-SR attacks. We use the OpenFace [134] face recognizer to evaluate the privacy protection performance of a probe face video. OpenFace extracts a 128-dimensional feature vector for each frame using a deep Convolutional Neural Network (CNN) and then uses a Support Vector Machine (SVM) classifier [135].

As privacy metric, we use cumulative rank-n identification accuracy η_i defined by Eq. 1.6. To measure fidelity, we use the PSNR that calculates the power ratio of the original frame with respect to the filtered frame in a video (see Eq. 1.1).

We measure flicker through subjective as well as objective evaluation. For the objective evaluation, we use the maximum of absolute difference ξ of pixel intensities defined by Eq. 1.3.

5.2.2 Parameter Selection

The value of α_t depends on the threshold ρ_j^o and the face movement. We evaluated the effect of α_t on the resulting flicker with different values of ρ_j^o on the detected faces in a video (Figure 5.11). The flicker of AHGMM^v depends on α_t and on ρ_j^o, with a higher variation at larger values of α_t. To achieve a flicker equal or less than AGB, α_t needs to be selected adaptively depending upon ρ_j^o, e.g. at $\rho_j^o = 0.6$ px/cm, an $\alpha_t \in \{0, 0.5\}$ is selected depending upon the face motion.



(a) (b)

Figure 5.11: Flicker ξ of (a) AGB at different thresholds ρ_j^o and of (b) AHGMM^v at different values of the smoothing factor α_t and ρ_j^o. ξ of AGB increases with ρ_j^o. In contrast, ξ of AHGMM^v depends on both α_t and ρ_j^o, and it negligibly increases with ρ_j^o, especially for low value of α_t; but significantly increases for high values of α_t.

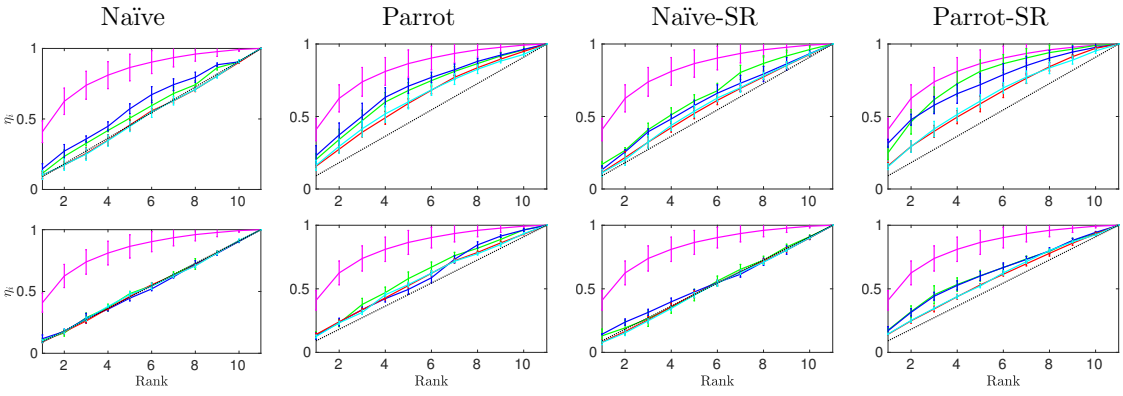


Figure 5.12: Rank-n identification accuracy, η_i, for privacy filters under naïve, parrot, naïve-SR and parrot-SR attacks at threshold $\rho_j^o = 0.6$ px/cm (first row) and $\rho_j^o = 0.4$ px/cm (second row). The filled marker shows the mean and the vertical bar the standard deviation of η_i for the multi-resolution frames. Legend: Unprotected, AHGMM^v, AHGMM^p, AGB, SVGB. AHGMM^v and AHGMM^p have the highest robustness against attacks (behaviour similar to a random classifier), especially under parrot and parrot-SR attacks.

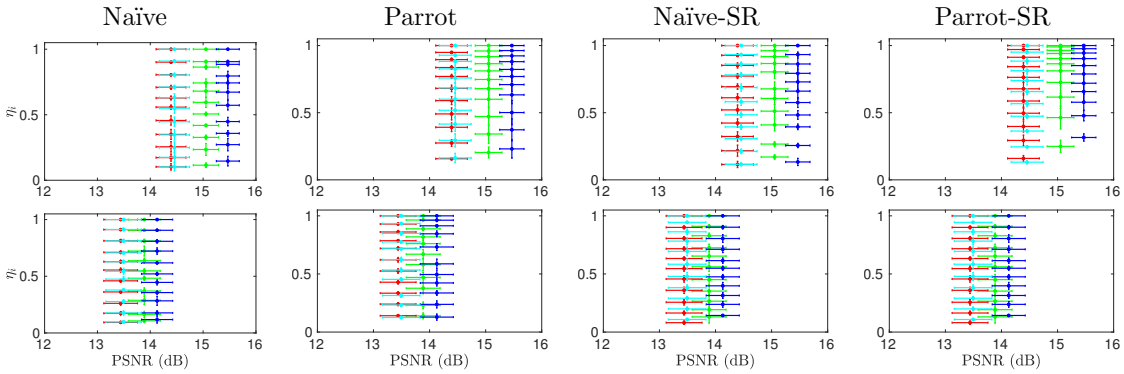


Figure 5.13: Relationship between rank-n identification accuracy, η_i, and fidelity, PSNR, for privacy filters under a naïve, parrot, naïve-SR and parrot-SR attacks at threshold $\rho_j^o =$ 0.6 px/cm (first row) and $\rho_j^o = 0.4$ px/cm (second row). The filled marker shows the mean of η_i and PSNR, the vertical bar indicates the standard deviation of η_i and the horizontal bar indicates the standard deviation of PSNR for the multi-resolution frames. Legend: AHGMM^v, AHGMM^p, AGB, SVGB. AHGMM^v leads to a slightly higher fidelity than AHGMM^p, due to temporal smoothing. SVGB uses the smallest Gaussian kernels for the outer parts of a face and leads to the highest value but with lower privacy protection.

5.2.3 Privacy Attacks

Figure 5.12 shows the results with the unprotected probe faces as the baseline and for a naïve, parrot, naïve-SR and parrot-SR attacks. Under the naïve attack, AHGMM^v and AHGMM^p maintain the highest privacy (i.e. η_i comparable to a random classifier) even with $\rho_j^o = 0.6$ px/cm where AGB and SVGB gives η_i larger than a random classifier. AHGMM^v achieves almost the same robustness as AHGMM^p against a parrot, naïve-SR and parrot-SR attack and it is unaffected by temporal smoothing. In contrast, AGB and SVGB filtered faces have lower privacy protection (i.e. larger η_i).

5.2.4 Fidelity analysis

Figure 5.13 shows the trade-off between η_i of a filter under different attacks and the corresponding fidelity. AHGMM^v has a slightly higher fidelity than AHGMM^p with almost similar values of η_i. This slightly increased fidelity is due to temporal smoothing which also

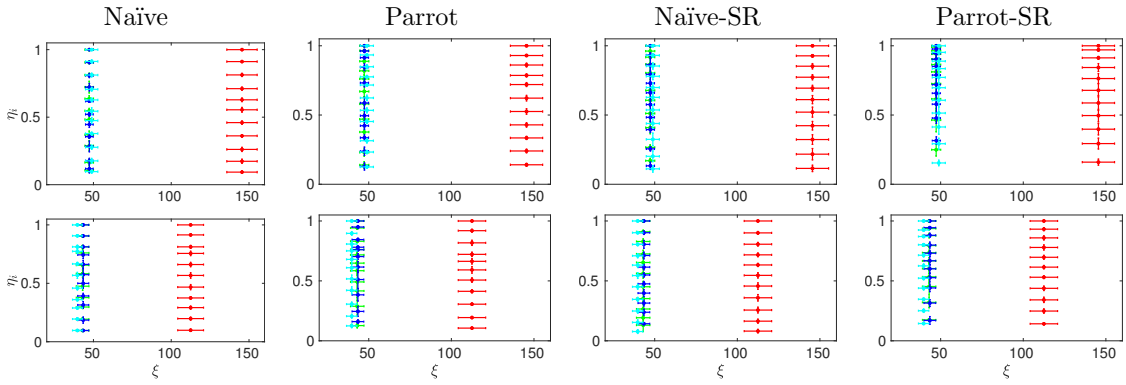


Figure 5.14: Relationship between rank-n identification accuracy, η_i, and flicker, ξ, for privacy filters under naïve, parrot, naïve-SR and parrot-SR attacks at threshold $\rho_j^o = 0.6$ px/cm (first row) and $\rho_j^o = 0.4$ px/cm (second row). The filled marker shows the mean of η_i and ξ, and the vertical and the horizontal bar indicate the standard deviation of η_i and ξ, respectively, for the multi-resolution frames. The larger ξ (see Eq. 1.3) the stronger the flicker. Legend: AHGMM^v, AHGMM^p, AGB, SVGB. AHGMM^v has a much lower than AHGMM^p and is similar to AGB and SVGB without any decrease in η_i, thus improving the trade-off between η_i and ξ.

minimise spatial distortion created by the switching kernels. In contrast, SVGB has the highest fidelity at the cost of lowest privacy (larger η_i), followed by AGB. The main reason of higher fidelity and lower privacy of SVGB compared to AGB is due to the outer parts of a face processed by smaller Gaussian kernels as SVGB linearly decrease the kernel size. In summary, AHGMM^v slightly improves fidelity while still robustly protecting the faces against different attacks.

5.2.5 Flicker Analysis

Objective Evaluation

Figure 5.14 depicts the trade-off between η_i of a filter under different attacks and the corresponding flicker measured using Eq. 1.3. The flicker generated by AHGMM^v is significantly lower than AHGMM^p with almost the same values of η_i for any threshold ρ_j^o. This lower flicker is the result of averaging of the frames with decaying weights that temporally reduce the effect of pseudo-random switching of the Gaussian kernels. In contrast, although SVGB

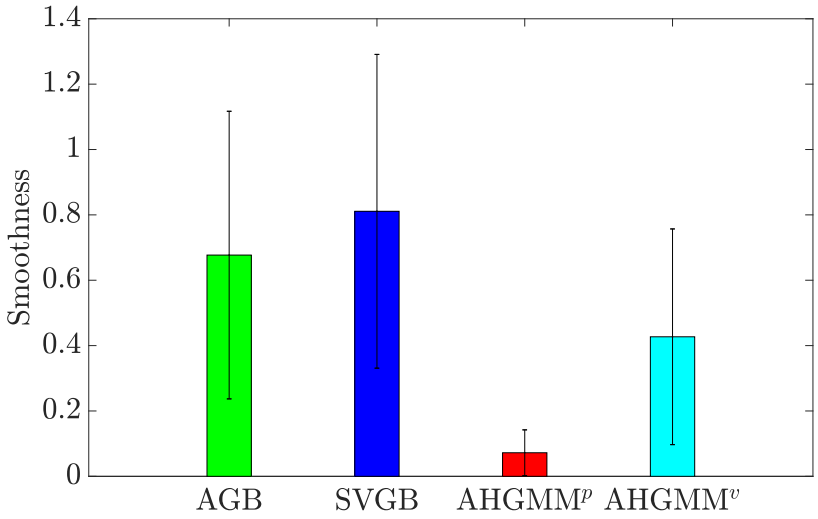


Figure 5.15: Subjective evaluation results. The bars indicate the mean (the larger the value, the more frequently videos processed by a method were chosen because of a better smoothness); the vertical lines represent the standard deviation.

and AGB have flicker similar to AHGMM^v, SVGB and AGB lead to η_i values (i.e. lower privacy). Comparatively, the flicker of SVGB is slightly greater than that of AGB, due to linear variation of the Gaussian kernels. AHGMM^v lowers flicker while being robust against the navïe, parrot, navïe-SR and parrot-SR attacks.

Subjective Evaluation

We finally evaluate flicker with a set of 20 human observers: 14 males and 6 females, aged between 25-35 years and without any image or video processing experience. We selected three videos captured with three different pitch angles/scales, and filter them with the four privacy filters: AGB, SVGB, AHGMM^p, and AHGMM^v (see Figure 5.16 and Figure 5.17). We paired the four filtered versions of each video, thus generating six combinations. The observers were asked to select the smoother video for each pair. Figure 5.15 shows the results of this subjective evaluation: AHGMM^v improves smoothness compared to AHGMM^p, but is less smooth than AGB and SVGB. This may be caused by small jerking in case of significant face movements and by the adaptation of α_t for maintaining the dynamics of the motion of the face.

Figure 5.16: Sample frames from the videos used in the subjective evaluation.

(a) (b) (c) (d) (e)

Figure 5.17: Face regions used in the subjective evaluation. (a) unprotected regions. Regions filtered with (b) AGB, (c) SVGB, (d) AHGMM^p and (e) AHGMM^v.

5.3 Summary

This chapter evaluated the validity of the AHGMM for both photography and videography scenarios using a state-of-the-art face recognition algorithm. For the photography scenario, the chapter used a synthetic face data set with faces at different pitch angles and resolutions emulating faces as captured from an MAV, while for the photography scenario it exploited a small real data set. It is found that the proposed algorithm provides the highest privacy under a parrot attack, an inverse-filter attack and a super-resolution attack. Moreover, a temporal smoother does not affect robustness against various attacks. This smoothness constraint slightly increases fidelity and significantly decreases the flicker.

Unlike face-de-identification approaches ([12, 40, 41, 44, 91, 92, 95]), the AHGMM does not depend on a sophisticated detector (i.e. pose, facial expression, age, gender, race) to counter parrot, inverse-filter and super-resolution attacks. Moreover, unlike the encryption/scrambling filters ([19, 20, 30, 31, 48, 73, 74, 76, 83–85]), the AHGMM prevents the recovery of the original face even with access to the seed of the PRNG.

Chapter 6

Conclusion

This chapter concludes the thesis by presenting its key contributions and limitations. Moreover, it discusses directions of future research areas.

6.1 Summary

In this thesis, the following contributions in the field of visual privacy protection are presented:

First, the thesis critically reviewed privacy filters intended for both fixed and mobile cameras. Specifically, it stated requirements of a privacy filter (i.e. robustness against different attacks, minimal spatio-temporal distortion and computational simplicity) for airborne recreational videography. Furthermore, it highlighted the limitations of the existing state-of-the-art privacy filters.

Second, the thesis presented the concept of a privacy design space for adaptive privacy filtering to minimise spatial distortion. Although any adaptive privacy filter (e.g. pixelation, Gaussian blur, scrambling, warping and morphing) could exploit it, evaluation using a Gaussian blur was performed that showed the fidelity improvement while exhibiting the same privacy level as provided by a fixed filter (naïve attack only).

Third, the thesis presented a novel privacy filter AHGMM for recreational photography that showed robustness not only against a naïve attack but also a brute force attack, a parrot attack and a reconstruction attack. Also, it minimised spatial distortion compared to a fixed filter.

Fourth, the thesis updated the AHGMM for videography by concatenating a temporal low

pass filter to minimise flicker. Experimental results showed that the temporal low pass filter significantly reduced flicker without losing robustness against different attacks.
Fifth, the thesis presented a synthetically generated face image data set of 480,000 faces belonging to 4281 subjects, emulating faces captured from 40 different positions, i.e. change of resolution and pitch angles. Moreover, the thesis also presented a small UHD face video data set of 11 subjects with a pitch angle variation of 20° - 78° and a height variation of 4 m - 7 m.

6.2 Future Directions

6.2.1 Limitations

The main limitation of both the privacy design space for adaptive filtering and the proposed AHGMM filter is their dependency on an experimentally determined threshold which needs to be calculated for each main-identifier using its state-of-the-art recognition algorithm. In future, if there is a more powerful recognition algorithm, then an update to the threshold is required. It is not known yet how to make the AHGMM or other adaptive privacy filters independent of such a threshold, but it is an important aspect to focus in future.
The second limitation of both the privacy design space for adaptive filtering and the proposed AHGMM filter is the assumption that the navigation sensors' data with high precision is available. In the thesis, this data was analytically calculated without using the actual navigation sensors. Thus, in future, it is important to investigate the effect of using actual navigation sensors' data, instead of analytical one by capturing a multi-modal face data set.
Finally, it is assumed that detection of the main-identifiers in each frame is given which may not be true practically due to wrong-detections or miss-detections. One possible way to minimise such wrong-detections or miss-detections could be to exploit the navigation sensors' data. For this purpose, a captured image can be divided into different sub-regions each with a prior knowledge of pixel density estimated using the navigation sensors' data. Consequently, through the Bayesian relation, the probability of wrong-detections or miss-detections for each sub-region can be minimised. Possibly, it could also improve true-detections.

6.2.2 Non-facial Identity Protection

The thesis only focused on protecting the main-identifiers and specifically used faces for the evaluation of the proposed privacy filters. Although the proposed filter is equally applicable to the other main-identifiers, e.g. vehicle licence plates and glass windows of private houses, it is first required to benchmark (i.e. threshold) corresponding state-of-the-art algorithms to claim privacy protection. Moreover, as indicated in Chapter 2, identity may leak form quasi-identifiers, e.g. age, gender, race, clothes type, clothes colour, location and time of an individual, this is one of the future research areas. Specifically, either by exploiting the existing state-of-the-art privacy filters or working from scratch, processing pipelines should be devised to achieve an ideal privacy filter, i.e. a machine algorithm that conceals both main-identifiers and quasi-identifiers.

6.2.3 Contextual Integrity

Both pre-possessing and post-processing privacy filters only ensure privacy protection within a given context. However, if there is a movement of individuals from one context to another (i.e. more often in airborne cameras), then identity may leak from the sensitive-regions even protected by an ideal privacy filter. Let I_R and I_{NR} be the values of I in the restricted area (private homes) and non-restricted area (public place), respectively. Furthermore, let both I_R and I_{NR} be made close to zero by applying a privacy filter, e.g. blanking out the restricted area and applying an ideal privacy filter for a detected individual in the non-restricted area. Now, consider an individual in a non-restricted area, who after some time enters in a restricted area. This entrance increases his/her identity level as very few people are authorised to enter the restricted area. Identity level obtained from this transition can be associated with a sensitive information revealed in the non-restricted area and consequently, there will be a privacy loss. The same applies for exiting a restricted area. Thus, in order to protect privacy, some mechanism should be adopted to maintain contextual integrity during such transitions between different areas.

Appendix A

Appendix

A.1 Face Image Data Set

Figure A.1 shows sample images of the stages of the dataset generation pipeline. We fit a 3D Morphable Model (3DMM) [137] on an input image to detect 68 facial landmarks [138] and then iteratively fit a 3DMM to generate a 3D image representation[1]. As there may be only a few degrees pitch of the subject captured in the images (e.g. a person looking slightly downward or upward), we rotate the 3D image at 0° pitch by applying a geometric transformation computed from the estimated pose of the fitted 3DMM. This disturbs the image alignment of the original data set, so a realignment is required, which we perform after generating the pitch effect. The synthetic pitch angles start from 0° to 70° with a step size of 10° and project it back to generate a corresponding 2D image. In order to align this image so that the eyes and nose appear at the same place among the images belonging to the same pitch angle, we apply an affine transformation computed by detecting eyes and nose tip using Dlib library [139] such that the transformed face has a resolution of 96×96 pixels. As the detection accuracy of the eyes and nose decrease with increasing pitch angle, we generate a ground truth (location of eyes and nose tip) of the 0° pitch angle images and uses it for the higher pitch angle images.

Finally, to introduce different height effects for the 8 synthetically generated images, we down-sample them with a factor of 2, 4, 8 and 16 generating images of 48×48, 24×24,

[1] Among the 12000 images, the landmark detector [138] was unable to detect 68 facial landmarks on 74 images. Therefore, we were unable to fit a 3DMM and used the original 74 images in order to comply with the standard verification test script of the LFW data set.

(a) (b) (c) (d) (e)

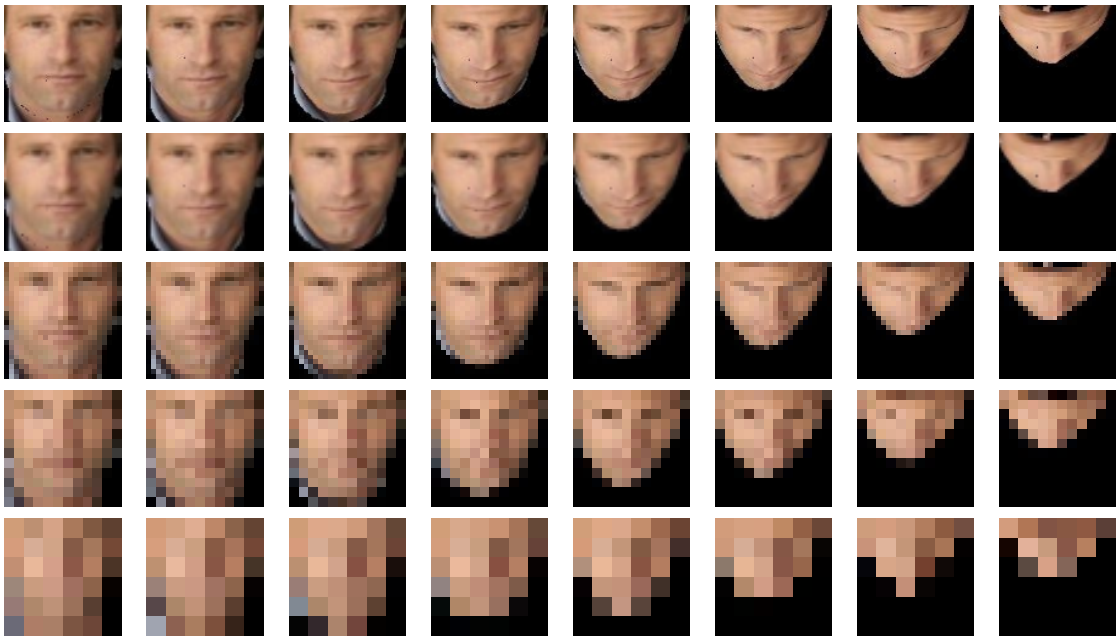
Figure A.1: Sample images at different stages during the data set generation process. (a) Original image 250 × 250 pixels, (b) image after fitting a 3D morphable model at 0° pitch angle, (c) image with synthetic pitch effect produced by applying a 3D geometric transformation, (d) aligned image of 96 × 96 pixels produced by applying an affine transformation computed by detecting eyes and nose location and (e) down-sampled image emulating an image captured at a different height.

12 × 12, 6 × 6 pixels, respectively. Thus, we increase the size of the original standard verification test of the LFW data set by 40 times, i.e. from 12000 images to 480,000 images. Fig. A.2 shows the 40 sample images belonging to the same and different subjects.
We manually determined the values of ρ_h and ρ_v by

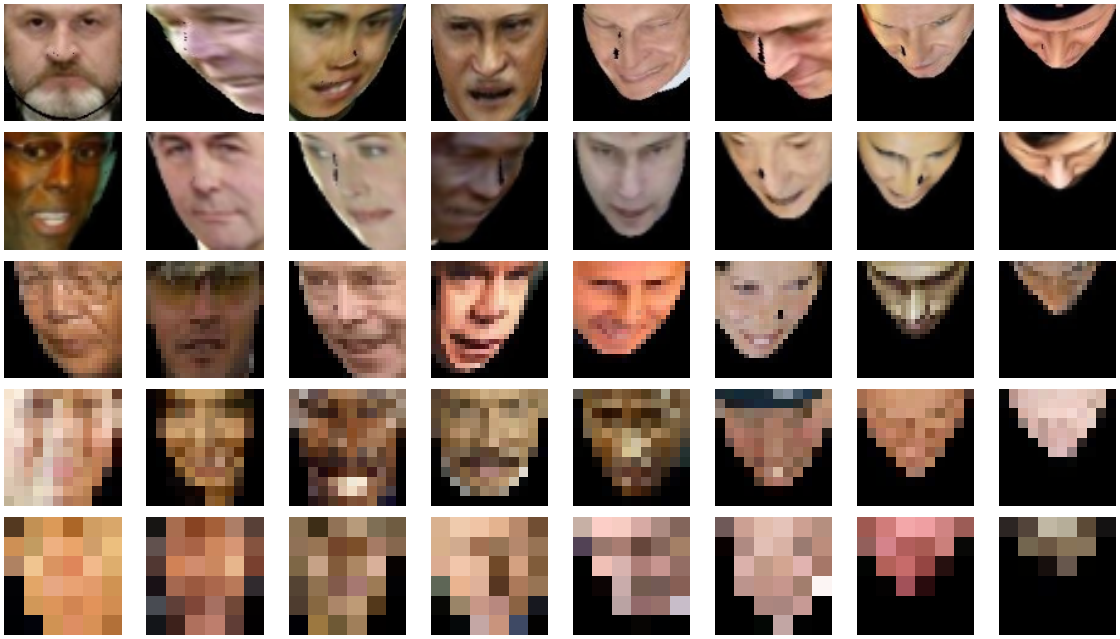
$$\rho_h = W/W_s, \tag{A.1}$$

$$\rho_v = W cos(\gamma)/H_s, \tag{A.2}$$

where W is the cropped width of a face in pixels, $\gamma = 90° - \theta_s$ is the pitch angle of the image and W_s and H_s are the average human face dimensions, i.e. the bitragion breadth of 15.45 cm and menton-crinion length of 20.75 cm, respectively [131].



(a)



(b)

Figure A.2: Sample images belonging to (a) a single subject and (b) multiple subjects from our synthetically generated airborne data set based on the LFW data set [133]. In each row, the pitch angle varies from 0° to 70° in 10° steps from left to right, while the image resolution remains same, i.e. first row: 96×96 pixels, second row: 48×48 pixels, third row: 24×24 pixels, fourth row: 12×12 pixels and fifth row: 6×6 pixels.

Bibliography

[1] S. Waharte and N. Trigoni. Supporting search and rescue operations with uavs. In *Proc. Int. Conf. on Emerging Security Technologies*, pages 142–147. Canterbury, UK, 2010.

[2] P. Tripicchio, M. Satler, G. Dabisias, E. Ruffaldi, and C. A. Avizzano. Towards smart farming and sustainable agriculture with drones. In *Proc. Int. Conf. on Intelligent Environments*, pages 140–143. Prague, Czech Republic, 2015.

[3] S. Park, L. Zhang, and S. Chakraborty. Design space exploration of drone infrastructure for large-scale delivery services. In *Proc. IEEE/ACM Int. Conf. on Computer-Aided Design*, pages 1–7. Austin, TX, USA, 2016.

[4] R. Babiceanu, P. Bojda, R. Seker, and M. Alghumgham. An onboard UAS visual privacy guard system. In *Proc. Integrated Communication, Navigation, and Surveillance Conf.*, pages J1:1–J1:8. Herdon, VA, USA, 2015.

[5] M. Sanfourche, B. L. Saux, A. Plyer, and G. L. Besnerais. Environment mapping amp; interpretation by drone. In *Proc. Joint Urban Remote Sensing Event*, pages 1–4. Lausanne, Switzerland, 2015.

[6] M. Quaritsch, K. Kruggl, D. Wischounig-Strucl, S. Bhattacharya, M. Shah, and B. Rinner. Networked uavs as aerial sensor network for disaster management applications. *e & i Elektrotechnik und Informationstechnik*, 127:56–63, 2010.

[7] Hexo+. https://hexoplus.com/. [Last accessed: 2018-10-27].

[8] AirDog. https://www.airdog.com/. [Last accessed: 2018-10-27].

[9] J. R. Padilla-López, A. A. Chaaraoui, and F. Flórez-Revuelta. Visual privacy protection methods: A survey. *Expert Systems with Applications*, 2015.

[10] M. Saini, P. Atrey, S. Mehrotra, and M. Kankanhalli. W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. *Multimedia Tools and Applications*, 68(1):135–158, 2014.

[11] J. Schiff, M. Meingast, D. Mulligan, S. Sastry, and K. Goldberg. Respectful cameras: detecting visual markers in real-time to address privacy concerns. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 971–978. San Diego, CA, USA, 2007.

[12] E. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying facial images. *IEEE Trans. on Knowledge and Data Engineering*, 17:232–243, 2005.

[13] P. Korshunov and T. Ebrahimi. Towards optimal distortion-based visual privacy filters. In *Proc. IEEE Int. Conf. on Image Processing*, pages 6051–6055. Paris, France, 2014.

[14] NoFlyZone. https://www.noflyzone.org/. [Last accessed: 2016-04-14].

[15] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y. L. Tian, A. Ekin, J. Connell, C. F. Shu, and M. Lu. Enabling video privacy through computer vision. *IEEE Security and Privacy*, 3(3):50–57, 2005.

[16] M. Saini, P. Atrey, S. Mehrotra, and M. Kankanhalli1. Anonymous surveillance. In *IEEE Int. Conf. on Multimedia and Expo*, pages 1–6. Barcelona, Spain, 2011.

[17] A. Cavallaro. Privacy in video surveillance. *IEEE Signal Processing Magazine*, 24(2):168–169, 2007.

[18] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, 1996.

[19] T. Boult. PICO: Privacy through Invertible cryptographic obscuration. In *Proc. Computer Vision for Interactive and Intelligent Environment*, pages 27–38. Lexington, KY, USA, 2005.

[20] F. Dufaux and T. Ebrahimi. Scrambling for privacy protection in video surveillance systems. *IEEE Trans. on Circuits and Systems for Video Technology*, 18(8):1168–1174, 2008.

[21] Á. Erdélyi, T. Barat, P. Valet, T. Winkler, and B. Rinner. Adaptive cartooning for privacy protection in camera networks. In *Proc. Int. Conf. on Advanced Video and Signal Based Surveillance*, pages 44–49. Seoul, Korea, 2014.

[22] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016.

[23] A. Cavallaro. Adding privacy constraints to video-based applications. In *Proc. European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, page 8. London, UK, 2004.

[24] S. Tansuriyavong and S. Hanaki. Privacy Protection by Concealing Persons in Circumstantial Video Image. In *Proc. Workshop on Perceptive User Interfaces*, pages 1–4. Orlando, FL, USA, 2001.

[25] D. Chen, Y. Chang, R. Yan, and J. Yang. Tools for protecting the privacy of specific individuals in video. *EURASIP Journal on Advances in Signal Processing*, 2007(1):107–107, 2007.

[26] K. Chinomi, N. Nitta, Y. Ito, and N. Babaguchi. Prisurv: Privacy protected video surveillance system using adaptive visual abstraction. In *Proc. Int. Conf. on Advances in Multimedia Modeling*, pages 144–154. Kyoto, Japan, 2008.

[27] F. Qureshi. Object-Video Streams for Preserving Privacy in Video Surveillance. In *Proc. Int. Conf. on Advanced Video and Signal Based Surveillance*, pages 442–447. Genova, Italy, 2009.

[28] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian. Privacy protecting data collection in media spaces. In *Proc. Int. Conf. on Multimedia*, pages 48–55. New York, NY, USA, 2004.

[29] O. Sarwar, B. Rinner, and A. Cavallaro. Design space exploration for adaptive privacy protection in airborne images. In *Proc. IEEE Advanced Video and Signal-based Surveillance*, pages 159–165. Colorado Springs, USA, 2016.

[30] P. Korshunov and T. Ebrahimi. Using warping for privacy protection in video surveillance. In *Proc. Int. Conf. on Digital Signal Processing*, pages 1–6. Fira, Santorini, Greece, 2013.

[31] F. Dufaux and T. Ebrahimi. Scrambling for video surveillance with privacy. In *Proc. Computer Vision and Pattern Recognition Workshops*, pages 160–160. New York, USA, 2006.

[32] J. X. Yang and H. R. Wu. A non-linear post filtering method for flicker reduction in H.264/AVC coded video sequences. In *Proc. IEEE Workshop on Multimedia Signal Processing*, pages 181–186. Cairns, Australia, 2008.

[33] S. Qiao, Y. Zhang, and H. Wang. PI-Frames for flickering reduction in H.264/AVC video coding. In *Proc. Int. Conf. on Computer Science and Service System*, pages 1551–1554. Nanjing, China, 2012.

[34] A. Jimnez-Moreno, E. Martnez-Enrquez, V. Kumar, and F. D. de Mara. Standard-compliant low-pass temporal filter to reduce the perceived flicker artifact. *IEEE Trans. on Multimedia*, 16(7):1863–1873, 2014.

[35] Z. Wen, J. Li, J. Liu, Y. Zhao, and J. Wen. Intra frame flicker reduction for parallelized HEVC encoding. In *Proc. Data Compression Conf.*, pages 111–120. Snowbird, UT, USA, 2016.

[36] S. B. Yoo, K. Choi, and J. B. Ra. Blind post-processing for ringing and mosquito artifact reduction in coded videos. *IEEE Trans. on Circuits and Systems for Video Technology*, 24(5):721–732, 2014.

[37] D. T. Vo, T. Q. Nguyen, S. Yea, and A. Vetro. Adaptive fuzzy filtering for artifact reduction in compressed images and videos. *IEEE Trans. on Image Processing*, 18(6):1166–1178, 2009.

[38] J. Yang, J. B. Park, and B. Jeon. Flickering effect reduction for H.264/AVC intra frames. *SPIE 6391, Multimedia Systems and Applications*, IX:6391–6399, 2006.

[39] Y. Kim, J. Jo, and S. Shrestha. A server-based real-time privacy protection scheme against video surveillance by unmanned aerial systems. In *Proc. Int. Conf. on Unmanned Aircraft Systems*, pages 684–691. Orlando, FL, USA, 2014.

[40] R. Gross, L. Sweeney, F. de la Torre, and S. Baker. Model-based face de-identification. In *Proc. Conf. on Computer Vision and Pattern Recognition Workshop*, pages 161–161. New York, USA, 2006.

[41] L. Du, M. Yi, E. Blasch, and H. Ling. Garp-face: Balancing privacy protection and utility preservation in face de-identification. In *Proc. IEEE Int. Joint Conf. on Biometrics*, pages 1–8. Clearwater, Florida, USA, 2014.

[42] Á. Erdélyi, T. Winkler, and B. Rinner. Privacy protection vs. utility in visual data. *Multimedia Tools and Applications*, pages 1–28, 2017.

[43] H. Yang, J. M. Boyce, and A. Stein. Effective flicker removal from periodic intra frames and accurate flicker measurement. In *Proc. IEEE Int. Conf. on Image Processing*, pages 2868–2871. San Diego, CA, USA, 2008.

[44] R. Gross, E. Airoldi, B. Malin, and L. Sweeney. Integrating utility into face de-identification. In *Proc. Privacy Enhancing Technologies Workshop*, pages 227–242. Cavtat, Croatia, 2005.

[45] Z. Sun, L. Meng, and A. Ariyaeeinia. Distinguishable de-identified faces. In *Proc. IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition*, pages 1–6. Ljubljana, Slovenia, 2015.

[46] B. Meden, . Emeri, V. truc, and P. Peer. K-Same-Net: k-Anonymity with generative deep neural networks for face de-identification. *Entropy*, 20(1), 2018.

[47] M. Saini, P. K. Atrey, S. Mehrotra, and M. Kankanhalli. Adaptive transformation for robust privacy protection in video surveillance. *Advances in Multimedia*, 2012:1–14, 2012.

[48] H. Sohn, D. N. Wesley, and M. R. Yong. Privacy protection in video surveillance systems: analysis of subband-adaptive scrambling in JPEG XR. *IEEE Trans. on Circuits and Systems for Video Technology*, 21(2):170–177, 2011.

[49] Y. Wang, M. ONeill, F. Kurugollu, and E. OSullivan. Privacy region protection for H.264/AVC with enhanced scrambling effect and a low bitrate overhead. *Signal Processing: Image Communication*, 35:71–84, 2015.

[50] P. C. Su, W. Y. Chen, S. Y. Shiau, C. Y. Wu, and A. Y. S. Su. A privacy protection scheme in H.264/AVC by data hiding. In *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–7. Kaohsiung, Taiwan, 2013.

[51] Z. Shahid, M. Chaumont, and W. Puech. Fast protection of H.264/AVC by selective encryption of CAVLC and CABAC for I and P Frames. *IEEE Trans. on Circuits and Systems for Video Technology*, 21(5):565–576, 2011.

[52] C. Dwork. Differential privacy. In *Proc. Int. Conf. on Automata, Languages and Programming - Volume Part II*, pages 1–12. Venice, Italy, 2006.

[53] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proc. Theory of Cryptography Conf.*, pages 265–284. New York, USA, 2006.

[54] J. Lee and C. Clifton. Differential identifiability. In *Proc. Int. Conf. on Knowledge Discovery and Data Mining*, pages 1041–1049. Beijing, China, 2012.

[55] H. Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79(1), 2004.

[56] T. Nawaz and J. Ferryman. An annotation-free method for evaluating privacy protection techniques in videos. In *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pages 1–6. Karlsruhe, Germany, 2015.

[57] T. Winkler and B. Rinner. Security and Privacy Protection in Visual Sensor Networks: A Survey. *ACM Compututing Surveys*, 47(1):2:1–2:42, 2014.

[58] Eagle Eye. *Bulletin of the Connecticut Academy of Science and Engineering*, 12(2), 1997.

[59] S. N. Patel, J. W. Summet, and K. N. Truong. *Protecting Privacy in Video Surveillance*, chapter BlindSpot: Creating Capture-Resistant Spaces, pages 185–201. Springer London, 2009.

[60] T. Yamada, S. Gohshi, and I. Echizen. Use of invisible noise signals to prevent privacy invasion through face recognition from camera images. In *Proc. ACM Int. Conf. on Multimedia*, pages 1315–1316. 2012.

[61] S. Zhu, C. Zhang, and X. Zhang. Automating visual privacy protection using a smart led. In *Proc. Int. Conf. on Mobile Computing and Networking*, pages 329–342. Snowbird, Utah, USA, 2017.

[62] Safe Haven. Safe haven from iceberg systems ensures privacy from camera phones; camera phone voyeurs and spy‘s can be defeated by new technology. http://www.m2.com/m2/web/story.php/20031E97F470D8BB9F6685256D9D006B597D. [Last accessed: 2018-10-27].

[63] V. Tiscareno, K. Jonhson, and C. Lawrence. Systems and methods for receiving infrared data with a camera designed to detect images. Google Patents, US 2011/0128384 A1, 2011.

[64] Y. Zhang, Y. Lu, H. Nagahara, and R. I. Taniguchi. Anonymous camera for privacy protection. In *Proc. Int. Conf. on Pattern Recognition*, pages 4170–4175. Stockholm, Sweden, 2014.

[65] F. Pittaluga and S. J. Koppal. Pre-capture privacy for small vision sensors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 39(11):2215–2226, 2017.

[66] J. Jung and M. Philipose. Courteous glass. In *Proc. ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1307–1312. Seattle, Washington, USA, 2014.

[67] J. Steil, M. Koelle, W. Heuten, S. Boll, and A. Bulling. PrivacEye: Privacy-Preserving First-Person Vision Using Image Features and Eye Movement Analysis. *ArXiv e-prints*, 2018.

[68] Federal Aviation Administration. B4ufly smartphone app. https://www.faa.gov/uas/where_to_fly/b4ufly/. [Last accessed: 2018-10-27].

[69] R. Allamaraju, H. Kingravi, A. Axelrod, G. Chowdhary, R. Grande, J. How, C. Crick, and W. Sheng. Human aware UAS path planning in urban environments using nonstationary MDPs. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 1161–1167. Hong Kong, China, 2014.

[70] G. Mcneal. Drones and aerial surveillance : Considerations for legislators. Technical report, Brookings Institution: The Robots Are Coming: The Project on Civilian Robotics; Pepperdine University Legal Studies Research Paper No. 2015/3, 2014.

[71] T. Vaidya and M. Sherr. Mind your (R, Φ)s: Location-Based Privacy Controls for Consumer Drones. In *Proc. Int. Workshop on Security Protocols*, pages 91–104. Cambridge, UK, 2015.

[72] L. Tang. Methods for encrypting and decrypting MPEG video data efficiently. In *Proc. ACM Int. Conf. on Multimedia*, page 219229. Boston, MA, USA, 1996.

[73] N. Baaziz, N. Lolo, O. Padilla, and F. Petngang. Security and privacy protection for automated video surveillance. In *Proc. IEEE Int. Symposium on Signal Processing and Information Technology*, pages 17–22. Cairo, Egypt, 2007.

[74] K. Yabuta, H. Kitazawa, and T. Tanaka. A new concept of security camera monitoring with privacy protection by masking moving objects. In *Proc. Advances in Multimedia Information Processing - PCM 2005*, pages 831–842. Jeju Island, Korea, 2005.

[75] P. Carrillo, H. Kalva, and S. Magliveras. Compression independent reversible encryption for privacy in video surveillance. *EURASIP Journal on Information Security*, pages 1–13, 2009.

[76] P. Korshunov and T. Ebrahimi. Using face morphing to protect privacy. In *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pages 208–213. Kraków, Poland, 2013.

[77] Y. Kusama, H. Kang, and K. Iwamura. Mosaic-based privacy-protection with reversible watermarking. In *Int. Joint Conf. on e-Business and Telecommunications*, volume 05, pages 98–103. Alsace, France, 2015.

[78] Y. Nakashima, T. Koyama, N. Yokoya, and N. Babaguchi. Facial expression preserving privacy protection using image melding. In *IEEE Int. Conf. on Multimedia and Expo*, pages 1–6. Torino, Italy, 2015.

[79] M. Bonetto, P. Korshunov, G. Ramponi, and T. Ebrahimi. Privacy in mini-drone based video surveillance. In *Workshop on De-identification for privacy protection in multimedia*, pages 1–6. Ljubljana, Slovenia, 2015.

[80] R. Jiang, S. Al-Maadeed, A. Bouridane, D. Crookes, and M. E. Celebi. Face recognition in the scrambled domain via salience-aware ensembles of many kernels. *IEEE Trans. on Information Forensics and Security*, 11(8):1807–1817, 2016.

[81] R. Jiang, A. Bouridane, D. Crookes, M. E. Celebi, and H. L. Wei. Privacy-protected facial biometric verification using fuzzy forest learning. *IEEE Trans. on Fuzzy Systems*, 24(4):779–790, 2016.

[82] N. Ruchaud and J. L. Dugelay. Aseppi: Robust privacy protection against de-anonymization attacks. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, pages 1352–1359. Honolulu, Hawaii, US, 2017.

[83] A. Chattopadhyay and T. Boult. Privacycam: a privacy preserving camera using uclinux on the blackfin dsp. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1–8. Minneapolis, MN, USA, 2007.

[84] M. Rahman, M. Hossain, H. Mouftah, A. El Saddik, and E. Okamoto. A real-time privacy-sensitive data hiding approach based on chaos cryptography. In *Proc. IEEE Int. Conf. on Multimedia and Expo*, pages 72–77. Suntec City, Singapore, 2010.

[85] T. Winkler and B. Rinner. Securing Embedded Smart Cameras with Trusted Computing. *EURASIP Journal on Wireless Communications and Networking*, 2011:8:1–8:20, 2011.

[86] X. Zhang, S. Seo, and C. Wang. A lightweight encryption method for privacy protection in surveillance videos. *IEEE Access*, 6:18074–18087, 2018.

[87] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent. Large-scale Privacy Protection in Google Street View. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 2373–2380. Kyoto, Japan, 2009.

[88] S. C. S. Cheung, J. Zhao, and M. V. Venkatesh. Efficient object-based video inpainting. In *Proc. Int. Conf. on Image Processing*, pages 705–708. Atlanta, GA USA, 2006.

[89] W. Zhang, S. S. Cheung, and M. Chen. Hiding privacy information in video surveillance system. In *Proc. IEEE Int. Conf. on Image Processing*, pages 868–871. Genova, Italy, 2005.

[90] M. Koelle, S. Ananthanarayan, S. Czupalla, W. Heuten, and S. Boll. Your smart glasses' camera bothers me!: Exploring opt-in and opt-out gestures for privacy mediation. In *Proc. Nordic Conf. on Human-Computer Interaction*, pages 473–481. Oslo, Norway, 2018.

[91] Y. Lin, S. Wang, Q. Lin, and F. Tang. Face swapping under Large Pose Variations: A 3D model based approach. In *Proc. IEEE Int. Conf. on Multimedia and Expo*, pages 333–338. Sydney, Australia, 2012.

[92] G. Letournel, A. Bugeau, V. T. Ta, and J. P. Domenger. Face de-identification with expressions preservation. In *Proc. IEEE Int. Conf. on Image Processing*, pages 4366–4370. Quebec city, Canada, 2015.

[93] B. Bhattarai, A. Mignon, F. Jurie, and T. Furon. Puzzling face verification algorithms for privacy protection. In *Proc. IEEE Int. Workshop on Information Forensics and Security*, pages 66–71. Atlanta, GA, USA, 2014.

[94] B. Driessen and M. Dürmuth. *Achieving Anonymity against Major Face Recognition Algorithms*, pages 18–33. Springer Berlin Heidelberg, 2013.

[95] P. Chriskos, O. Zoidi, A. Tefas, and I. Pitas. De-identifying facial images using projections on hyperspheres. In *Proc. IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition*, volume 04, pages 1–6. 2015.

[96] P. Chriskos, O. Zoidi, A. Tefas, and I. Pitas. De-identifying facial images using singular value decomposition and projections. *Multimedia Tools and Applications*, pages 1–34, 2016.

[97] K. Brki, I. Sikiri, T. Hrka, and Z. Kalafati. De-identifying people in videos using neural art. In *Int. Conf. on Image Processing Theory, Tools and Applications*, pages 1–6. Oulu, Finland, 2016.

[98] S. Cifti, A. O. Akyüz, and T. Ebrahimi. A reliable and reversible image privacy protection based on false colors. *IEEE Trans. on Multimedia*, 20(1):68–81, 2018.

[99] R. Templeman, M. Korayem, D. Crandall, and K. Apu. Placeavoider: Steering first-person cameras away from sensitive spaces. In *Proc. Network and Distributed System Security Symposium*, February, pages 23–26. San Diego, CA, USA, 2014.

[100] P. Aditya, R. Sen, P. Druschel, S. Joon Oh, R. Benenson, M. Fritz, B. Schiele, B. Bhattacharjee, and T. T. Wu. I-pic: A platform for privacy-compliant image capture. In *Proc. Int. Conf. on Mobile Systems, Applications, and Services*, pages 235–248. Singapore, Singapore, 2016.

[101] R. Hatem A., G. Miguel A., A. M. Ballest, and P. Domnec. Defeating face de-identification methods based on DCT-block scrambling. *Machine Vision and Applications*, 27:251262, 2015.

[102] H. Hofbauer, A. Unterweger, and A. Uhl. Encrypting only ac coefficient signs considered harmful. In *Int. Conf. on Image Processing*, pages 3740–3744. Quebec city, Canada, 2015.

[103] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen. Image Melding: Combining inconsistent images using patch-based synthesis. *ACM Trans. on Graphics*, 31(4):82:1–82:10, 2012.

[104] H.-J. Hsu and K.-T. Chen. Face Recognition on Drones: Issues and Limitations. In *Proc. First Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use*, DroNet '15, pages 39–44. Florence, Italy, 2015.

[105] R. McPherson, R. Shokri, and V. Shmatikov. Defeating image obfuscation with deep learning. *arXiv e-prints*, abs/1609.00408:1–12, 2016.

[106] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. *arXiv e-prints*, abs/1508.06576:1–16, 2015.

[107] S. Gao, J. Ma, W. Shi, G. Zhan, and C. Sun. Trpf: A trajectory privacy-preserving framework for participatory sensing. *IEEE Trans. on Information Forensics and Security*, 8(6):874–887, 2013.

[108] H. Kido, Y. Yanagisawa, and T. Satoh. An anonymous communication technique using dummies for location-based services. In *Proc. Int. Conf. on Pervasive Services*, pages 88–97. Santorini, Greece, 2005.

[109] L. Sweeney. K-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570, 2002.

[110] B. Gedik and L. Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Trans. on Mobile Computing*, 7(1):1–18, 2008.

[111] M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In *Proc. Int. Conf. on Pervasive Computing*, pages 152–170. Munich, Germany, 2005.

[112] C.-Y. Chow and M. F. Mokbel. Trajectory privacy in location-based services and data publication. *SIGKDD Explor. Newsl.*, 13(1):19–29, 2011.

[113] M. E. Nergiz, M. Atzori, and Y. Saygin. Towards trajectory anonymization: A generalization-based approach. In *Proc. ACM Int. Workshop on Security and Privacy in GIS and LBS*, pages 52–61. Irvine, CA, USA, 2008.

[114] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *Proc. IEEE Int. Conf. on Data Engineering*, pages 376–385. Washington, DC, USA, 2008.

[115] T. Xu and Y. Cai. Exploring historical location data for anonymity preservation in location-based services. In *Proc. IEEE INFOCOM 2008. The 27th Conf. on Computer Communications*. Phoenix, AZ, USA, 2008.

[116] T. Ahonen, A. Hadid, and M. Pietikainen. Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[117] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[118] British Security Industry Association (BSIA). *Planning, design, installation and operation of CCTV surveillance systems code of practice and associated guidance*, 2014.

[119] European Commmittee for Standardization (CEN). *Alarm systems CCTV surveillance systems for use in security applications, part 7, EN 50132-7*, 2012.

[120] Axis Communication. Perfect pixel count meeting your operational requirements. https://www.axis.com/files/feature_articles/ar_perfect_pixel_count_55971_en_1402_lo.pdf. [Last accessed: 2018-10-27].

[121] T. Marciniak, A. Chmielewska, R. Weychan, M. Parzych, and A. Dabrowski. Influence of low resolution of images on reliability of face detection and recognition. *Multimedia Tools and Applications*, 74(12):4329–4349, 2015.

[122] C. Lu and X. Tang. Surpassing human-level face verification performance on LFW with gaussian face. In *Proc. AAAI Conf. on Artificial Intelligence*, pages 3811–3819. 2015.

[123] O. Sarwar, B. Rinner, and A. Cavallaro. Concealing the identity of faces in oblique images with adaptive hopping Gaussian mixtures. *ArXiv e-prints*, abs/1810.12435:1–19, 2018.

[124] O. Sarwar, B. Rinner, and A. Cavallaro. Temporally smooth privacy protected airborne videos. In *Proc. IEEE Int. Conf. on Intelligent Robots*, pages 1–6. Madrid, Spain, 2018.

[125] D. Turner, A. Lucieer, and L. Wallace. Direct Georeferencing of Ultrahigh-Resolution UAV Imagery. *IEEE Trans. on Geoscience and Remote Sensing*, 52(5):2738–2745, 2014.

[126] J. Höhle. Oblique aerial images and their use in cultural heritage documentation. In *Proc. Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 349–354. Strasbourg, France, 2013.

[127] P. Meixner and F. Leberl. Interpreting building facades from vertical aerial images using the third dimension. In *Proc. Joint Symposium of ISPRS Technical Commission IV & AutoCarto*, pages 15–19. Orlando, FL, USA, 2010.

[128] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals & Systems (2nd Ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[129] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on Image Processing*, 13(4):600–612, 2004.

[130] D. L. Baggio, S. Emami, D. M. Escriva, K. Ievgen, N. Mahmood, J. Saragih, and R. Shilkrot. *Mastering OpenCV with Practical Computer Vision Projects*. Packt Publishing, Limited, 2012.

[131] Human Engineering Design Data Digest, Department of Defense Human Factors Engineering Technical Advisory Group (DOD HFE TAG). http://www.acq.osd.mil/rd/hptb/hfetag/products/documents/HE_Design_Data_Digest.pdf, 2000. [Last accessed: 2018-10-27].

[132] T. Popkin, A. Cavallaro, and D. Hands. Accurate and efficient method for smoothly space-variant gaussian blurring. *IEEE Trans. on Image Processing*, 19(5):1362–1370, 2010.

[133] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[134] B. Amos, B. Ludwiczuk, and M. Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.

[135] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 815–823. Boston, MA, USA, 2015.

[136] J. L. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia. Diatom autofocusing in brightfield microscopy: a comparative study. In *Proc. Int. Conf. on Pattern Recognition*, pages 314–317. Barcelona, Spain, 2000.

[137] A. Bas, W. A. P. Smith, T. Bolkart, and S. Wuhrer. Fitting a 3D morphable model to edges: A comparison between hard and soft correspondences. In *Proc. Asian Conf. on Computer Vision*, pages 1–15. Taipei, Taiwan, 2016.

[138] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2879–2886. Providence, RI, USA, 2012.

[139] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.